12-13-2019

Information retrieval from spaceborne GNSS Reflectometry observations using physics- and learning-based techniques

Orhan Eroglu

Information retrieval from spaceborne GNSS Reflectometry observations using physics- and learning-based techniques

By

Orhan Eroglu

A Dissertation
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
in Electrical and Computer Engineering
in the Department of Electrical and Computer Engineering

Mississippi State, Mississippi

December 2019

Information retrieval from spaceborne GNSS Reflectometry

observations using physics- and

learning-based techniques

By

Orhan Eroglu

Approved:

Mehmet Kurum
(Major Professor)

John E. Ball
(Committee Member)

J. Patrick Donohoe
(Committee Member)

Ali Gurbuz
(Committee Member)

Qian (Jenny) Du
(Interim Graduate Program Director)

Jason M. Keith
Dean
Bagley College of Engineering

Name: Orhan Eroglu

Date of Degree: December 13, 2019

Institution: Mississippi State University

Major Field: Electrical and Computer Engineering

Major Professor: Mehmet Kurum

Title of Study: Information retrieval from spaceborne GNSS Reflectometry observations using physics- and learning-based techniques

Pages of Study: 250

Candidate for Degree of Doctor of Philosophy

This dissertation proposes a learning-based, physics-aware soil moisture (SM) retrieval algorithm for NASA's Cyclone Global Navigation Satellite System (CYGNSS) mission. The proposed methodology has been built upon the literature review, analyses, and findings from a number of published studies throughout the dissertation research. Namely, a Signals of Opportunity Coherent Bistatic scattering model (SCoBi) has been first developed at MSU and then its simulator has been open-sourced. Simulated GNSS-Reflectometry (GNSS-R) analyses have been conducted by using SCoBi. Significant findings have been noted such that (1) Although the dominance of either the coherent reflections or incoherent scattering over land is a debate, we demonstrated that coherent reflections are stronger for flat and smooth surfaces covered by low-to-moderate vegetation canopy; (2) The influence of several land geophysical parameters such as SM, vegetation water content (VWC), and surface roughness on the bistatic reflectivity was quantified, the dynamic ranges of reflectivity changes due to SM and VWC are much higher than the changes due to the

surface roughness. Such findings of these analyses, combined with a comprehensive literature survey, have led to the present inversion algorithm: Physics- and learning-based retrieval of soil moisture information from space-borne GNSS-R measurements that are taken by NASA's CYGNSS mission. The study is the first work that proposes a machine learning-based, non-parametric, and non-linear regression algorithm for CYGNSS-based soil moisture estimation. The results over point-scale soil moisture observations demonstrate promising performance for applicability to large scales. Potential future work will be extension of the methodology to global scales by training the model with larger and diverse data sets.

Key words: bistatic scattering, CYGNSS, GNSS-Reflectometry, information retrieval, learning-based, physics-aware, SCoBi, Signals of Opportunity, soil moisture

DEDICATION

To my dearests and true accomplishments, Alime, Osman, and Hakan.

ACKNOWLEDGEMENTS

This work was supported in part by grant 80NSSC18K1329 from the National Aeronautics and Space Administration (NASA) Earth and Space Science Fellowship Program. The findings and opinions in this dissertation belong solely to the author, and are not necessarily those of the sponsor.

I would like to acknowledge Dr. Mehmet Kurum's exceptional supervision; without which I could not have produced this work. He always had time to discuss, share, and help without an exception.

I would like to thank Dr. John E. Ball and Dr. Ali Gurbuz for their invaluable support to my research. I also thank my committee for their reviews and comments on this dissertation.

The friendship of my colleagues Dylan R. Boyd and Robiul Hossain MD. Rafi as well as the support of The Electrical and Computer Engineering Department of Mississippi State University cannot go unnoticed.

Peer reviews at the CYGNSS Science Team Meetings have put many contributions to this research.

Final words go to my lifelong supporters and true fortunes: My parents Ihsan and Medine, my siblings Sukran, Sinan, Dudu, and Mustafa, and my nephews Yasin, Ece, Yagmur, Esin, and Elif.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Microwave remote sensing has been exploited for monitoring of the Earth surface (ocean, land, ice, snow, and vegetation) at global scales for decades. A few examples of applications can be listed as the sensing of ocean surface winds, altimeter, surface soil moisture (SM), sea-ice salinity, snow water equivalent, and vegetation water content [160, 178]. Traditional monostatic radars and radiometers have been commonly used for these applications. In addition, a third approach named Signals of Opportunity (SoOp) has been emerging in recent years. This dissertation investigates the Global Navigation Satellite System Reflectometry (GNSS-R) as an application domain of SoOp for surface SM retrieval because GNSS-R has a great potential to complement traditional instruments for improved spatio-temporal resolutions at global scales.

The rest of this short introductory chapter is organized as follows: Section 1.1 gives information about the importance of global SM monitoring and how it has been performed so far, followed by Section 1.2 with the motivation of this Ph.D. research for remote sensing of SM at global scales by using space-borne GNSS-R signals. Section 1.3 provides an overview of the organization of the dissertation. Section 1.4 concludes the summary of this dissertation's contributions to GNSS-R based SM retrieval.

1.1 Background and Literature Review

SM is a key land geophysical parameter on the continuous water, energy, and carbon fluxes through the ground-air boundary of the Earth because it is a major component governing the high energy-demanding evapotranspiration processes. More precisely, SM has a direct effect in the never-ending cycle of 10% of the atmospheric humidity. Furthermore, the thermal conductivity of the soil is driven by its moisture content, which in turn affects the regulation of the temperature over the surface [38]. Due to these facts, SM is considered within the Essential Climate Variables for improved climate knowledge [59]. Understanding SM at global scales with high spatio-temporal resolutions can help improve applications such as climate modeling, weather forecast, crop yield estimation, as well as natural hazard detection and prediction (e.g. flood, landslide, and drought) [76, 121, 41, 103].

Global SM retrieval has been performed by microwave remote sensing techniques with the use of passive and active instruments so far. For the former, radiometers have been commonly employed. These passive instruments measure the surface emissivity by sensing the brightness temperature. Surface SM can then be retrieved from emissivity since the measured emissivity is driven by the soil dielectric constant, observation angle, vegetation canopy, and surface roughness [76, 168, 160, 161]. The sensitivity of radiometers to SM has been demonstrated by several studies in the past [75, 173, 37]. These instruments, on the other hand, require to observe large fields to detect enough energy from the low quantities of microwave emissions over the surface; this leads to coarse spatial resolutions (several tens of kilometers) [7].

The L-band frequencies (1 to 2 GHz) are of particular interest for SM retrieval radiometry because signals within this spectrum are highly sensitive to the dielectric properties of the top 0-5 cm of the soil that varies with changing moisture content [122, 35, 161]. Moreover, these signals can penetrate well through cloud covers and non-heavy vegetation canopy [170, 122, 32], which are significant limitations for optical sensing techniques. Even within the microwave spectrum, higher frequencies than L-band (hence shorter wavelength) would introduce much more attenuation due to the atmosphere and vegetation. In addition, L-band radiometers have a reduced sensitivity to surface roughness [14]. European Space Agency's (ESA) Soil Moisture and Ocean Salinity (SMOS) [86] and National Aeronautics and Space Administration's (NASA) Soil Moisture Active Passive (SMAP) [42] satellites are current space-borne missions that have on-board L-band radiometers. Both of these missions provide SM products with a spatial resolution of around 36 km and revisit time of 2-3 days.

Radar backscattering measurements have been also used for global SM retrieval. This technique relies on the extraction of information from the own illumination of active instruments (radars) in a monostatic scheme. In other words, these instruments transmit electromagnetic signals to the Earth surface, and the backscattered signals are received by the same instruments. Radars measure the intensity of the backscattering coefficient by measuring the difference between the transmitted and received signal powers. SM can be estimated from backscattering measurements by determining the soil dielectric constant because there is a direct link between the observed backscattering coefficient and the dielectric constant of the scattering medium [7]. A number of previous studies have

successfully demonstrated the sensitivity of radars to SM [158, 159, 73, 34]. Radars can generally provide finer spatial resolutions (several tens of meters to few kilometers) with increased revisit times (few days to few weeks) compared to radiometers [7].

The global estimation of SM by using radar instruments has been performed by multiple space-borne missions. For instance, ESA's Advanced Synthetic Aperture Radar (ASAR) instrument (C-band) on-board the Environmental Satellite (Envisat) platform was launched in 2002 with increased coverage and spatial resolution capabilities for SM retrieval compared to its predecessors [145, 125, 127]. SMAP's monostatic radar that operated at L-band provided 3-km SM products with a revisit time of 2-3 days until its failure in mid 2015 [42]. In addition, backscattering measurements from ESA's Sentinel-1 (C-band) [126], Advanced Scatterometer (ASCAT) (C-band) [169], JAPAN's Phased Array L-band Synthetic Aperture Radar (PALSAR) [137], and DLR's TERRASAR-X (X-band) [6] were used for global SM estimates.

Within the recent decades, a new microwave remote sensing approach, SoOp, has emerged. SoOp is based on reception and assessment of the free-transmitted signals in a bistatic geometry, where such signals reflect off the Earth's surface in a forward direction to the SoOp receivers [57, 15]. From this perspective, the SoOp approach offers cost-effective global monitoring potential, requiring only the installation of a passive receiver and no need for a transmitter. Therefore, the research and application interest to use such reflected signals for the remote sensing of bio-geophysical parameters has been increasing.

GNSS-R is a particular domain of SoOp, which has been investigated for remote sensing of Earth since 1990's. In its application, free transmitters are the existing GNSS satel-

lites. GNSS-R configurations (GNSS-R receivers and free GNSS transmitters) function as a bistatic radar where the transmitter is not under the receiver side's control [178].

An illustration of a general GNSS-R bistatic geometry can be seen in Figure 1.1. This figure shows the relative movement of each platform (i.e. the GNSS transmitter and GNSS-R receiver) with the velocity vectors V_r and V_t, as well as the shortest multipath distance reflection point (i.e. specular point) with the point SP. The x-y coordinate defines the Earth's surface. GNSS-R receivers process their signal power measurements as delay-Doppler maps (DDM) [179]. The temporal differences between arrival of the direct signals (signals that are received through the direct path between the GNSS satellite and GNSS-R receiver) and reflected signals (signals that are received from the scattering over the Earth's surface) result in varying time delays (τ_i). The frequency differences between the direct signals and the reflected signals due to the relative speeds of the platforms cause Doppler shifts (f_i). There will be points on the scattering surface that will have equal delay or Doppler levels. For instance, the elliptical lines on Figure 1.1 show iso-Delay lines, whereas gray lines show the iso-Doppler lines. DDM thus consists of delay-Doppler bins that show the signal powers received from varying delay and Doppler values around SP. On the other hand, some points such as P and Q would have the same delay and Doppler values which results in an ambiguity that their corresponding DDM bin might be mapping to either of these points. The red line, however, defines an ambiguity-free region over SP where no two points can have the same delay and Doppler values [138].

The large number of satellites in any GNSS constellation (such as GPS, GLONASS, Galileo, and BeiDou) in the medium Earth orbit allows for frequent revisit times of the

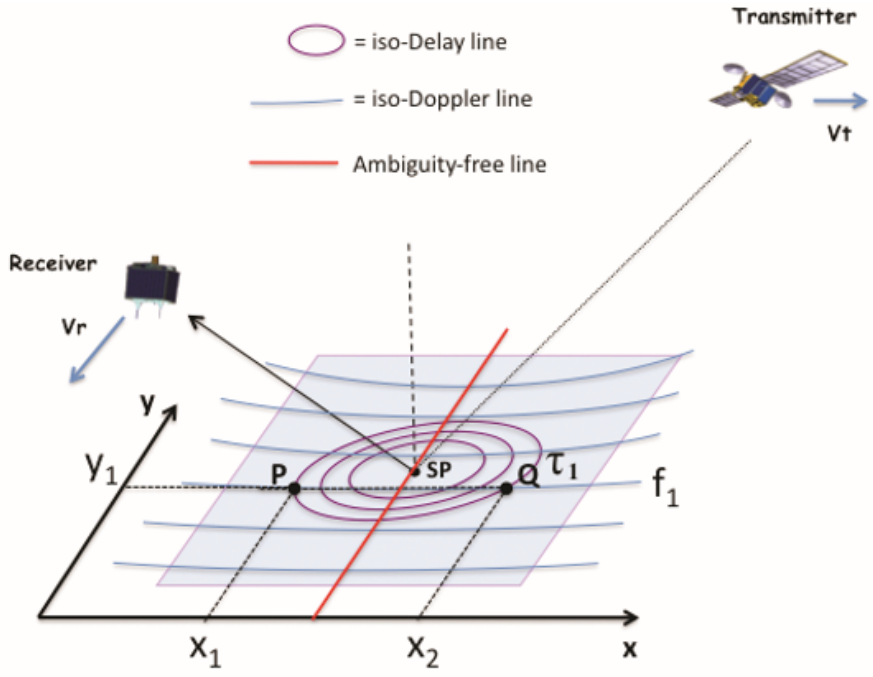

Figure 1.1: Bistatic geometry [138]

GNSS-R specular points [70]. Furthermore, several small-form factor GNSS-R satellites can be launched in a GNSS-R constellation as well to further improve revisit times. Thus, the GNSS-R technique has the potential to provide global coverage at high spatio-temporal scales by using small SWaP-C (Size Weight and Power - Cost) receivers in micro satellites [178].

GNSS-R capabilities have been investigated for several remote sensing applications by using either ground-based or airborne GNSS-R instruments with varying configurations. These applications include ocean altimetry [111], ocean surface roughness and wind retrieval [107, 179], sea-ice sensing [90], SM estimation [101, 97, 134, 39, 4, 40, 17, 25], snow depth sensing [98], as well as biomass estimation and vegetation growth monitoring [40, 153].

Several space-borne missions have been established to conduct GNSS-R observations at a global basis as well. For instance, the first dedicated space-borne GNSS-R receiver was a secondary payload on-board the United Kingdom Disaster Monitoring Constellation (UK-DMC) [61]. It has demonstrated the potentiality of GNSS-R for the remote sensing of ocean, ice, and land geophysical parameters from space [60]. The UK Technology Demonstration Satellite (TDS-1) was launched in 2014 with an improved secondary GNSS-R payload to demonstrate SoOp's ocean remote sensing feasibility [54]. NASA's Cyclone Global Navigation Satellite System (CYGNSS) is a GNSS-R mission with its eight small satellites each with four channels in orbit. The constellation was launched in December, 2016 to improve weather predictions by estimating ocean winds between 38° south and 38° north latitudes [143, 141]. CYGNSS has a mean revisit time of seven hours over the ocean, thanks to the multiple satellites. Publicly available wind speed retrieval data of the constellation demonstrated the maturity of the GNSS-R technique for ocean applications [142, 144, 29, 140].

Utilization of the GNSS-R methodology for the global sensing of SM, too, has been an emerging research interest of the Earth science community recently. A number of motivating factors can be listed for this as follows: **(i)** GNSS-R receivers exhibit similar sensitivity to SM as well as reduced attenuation due to clouds and non-heavy vegetation covers as in the L-band radiometers since they receive the L-band GNSS signals [170, 122, 32]. **(ii)** GNSS-R demonstrates potentiality to complement the spatial resolutions of the traditional microwave remote sensing techniques. **(iii)** Several small satellites can be launched in

GNSS-R constellations, which in turn could improve the revisit times. **(iv)** Small SWaP-C nature of the GNSS-R receivers can help reduce the operational costs.

Surface reflectivity can be derived from the GNSS-R observations by using the bistatic received power. Surface SM can subsequently be retrieved from the reflectivity since the measured reflectivity is driven by the soil dielectric constant, observation angle, and surface roughness [178]. From this point, GNSS-R is considered as analogous to L-band radiometry regarding its sensitivity to SM and surface roughness because conservation of energy requires a link between emissivity and reflectivity [112, 5, 51].

Past studies that investigated the effect of SM on GNSS-R observables can be grouped into four main categories: **(i)** Modeling and simulation studies, **(ii)** Ground-based and air-borne experiments, **(iii)** Qualitative analyses for space-borne observations, and **(iv)** Quantitative analyses and proposed SM retrieval methodologies. The first category, modeling and simulation efforts, relies on the use of forward scattering models, which are developed for unveiling the electromagnetic and physical mechanisms through the bistatic scattering processes. Previous studies using such models were applied to a range of microwave spectra including L-band (as in the GNSS-R case) under various bistatic geometries and land covers such as bare soil, croplands, and tree canopies [106, 52, 66, 130, 174]. Despite the fact that these studies were not directly aimed at providing GNSS-R based SM retrieval, they are worth being investigated for development of inversion models since they can account for the impact of SM on the GNSS-R deliverables.

A number of ground-based and air-borne experiments were conducted to study the applicability of GNSS-R to SM sensing. The GNSS interferometric reflectometry (GNSS-IR)

experiments held a reasonable fraction of ground-based GNSS-R efforts [134, 4, 39, 97, 25]. In these experiments, interference patterns between the direct and reflected signals, which are received by ground-based geodetic or specifically designed GNSS-IR receivers, are observed in order to perform SM retrieval. Findings of these experiments resulted in a project called PBO-H2O funded by National Science Foundation (NSF). The PBO-H2O network of multiple GNSS-IR instruments provides *in situ* SM products that are archived at the International Soil Moisture Network (ISMN) data base [99, 100, 152]. Other ground-based experiments were also conducted to investigate GNSS-R applicability for remote sensing of SM by using on-tower GNSS-R receivers. For example, Egido et al. [39] performed a long term experiment to observe GNSS-R polarimetric measurements over a crop field from a ground-based stationary platform for different soil moisture conditions. Airborne polarimetric experiments were conducted to investigate the ground-reflected GNSS signals' sensitivity to SM over larger scales [114, 113, 82, 12, 40].

With the launch of several GNSS-R space missions in 2000's, analysis of space-borne GNSS-R observatory data has emerged. Qualitative investigations were conducted to demonstrate GNSS-R sensitivity to SM from space since none of these missions were originally designed for SM retrieval, and applicability of space-borne GNSS-R signals for remote sensing of SM has been an open issue for years. Such studies typically try to unveil accompaniment between the spatial and/or temporal changes in GNSS-R measurements and reference SM values. For instance, UK's TDS-1 land observations were compared to ESA's SMOS SM products to examine the correspondence between reflected GNSS signals and surface SM under various vegetation covers [13, 23]. Some other studies have

also demonstrated the sensitivity of NASA's CYGNSS measurements to SM in a similar approach [118, 180, 142, 19]. One of the major contributions of such qualitative analyses to GNSS-R based SM retrieval research is that the demonstrated sensitivity of space-borne GNSS-R observations to SM motivated researchers to further investigate various SM estimation methodologies.

There are a number of studies that proposed GNSS-R based SM retrieval in conjunction with extensive quantitative analyses. Because the overall appraoch and findings of such studies are related to the motivation of this dissertation, they will be detailed in the next section.

1.2 Motivation

The CYGNSS constellation records continuous land measurements between 38° south and 38° north latitudes. These measurements are publicly available and construct a considerably large GNSS-R data set over land that can be analyzed. This increased size of satellite data helped researchers extensively investigate space-borne GNSS-R observations' applicability for SM retrieval in practice [24, 28, 87, 2]. Most of these studies proposed CYGNSS-based SM retrieval approaches that are generally based on correlating CYGNSS observations to the current state-of-the-science SM products from space-borne missions such as NASA's Soil Moisture Active Passive (SMAP) [42] and ESA's Soil Moisture and Ocean Salinity (SMOS) [86]. Hence, they compared their findings to the coarse spatial resolution (around 36 km) of these missions.

Despite the increasing number of studies on the CYGNSS-based SM retrieval, there are still open issues and challenges in the development of an operational GNSS-R based SM retrieval methodology. These will be extensively investigated throughout the rest of this dissertation; however, the most significant ones that attracted this Ph.D. research can be described as follows: **(i)** Bistatic received power would be dominated by either coherent reflections around a small neighborhood of the SP, or incoherent scattering over a much larger area, depending on a number of bio/geophysical factors such as topography, vegetation cover, and surface roughness [178]. Determination of whether the signal is coherent or incoherent is not an easy task, regarding the spatial variability of these factors over land. **(ii)** The vegetation attenuation, surface roughness degradation, as well as the observation angle effects require careful handling (i.e. corrections to the GNSS-R observables) in either case, too. **(iii)** Bistatic measurements are non-linearly dependent on the aforementioned bio/geophysical parameters; hence, the SM retrieval process requires advance non-linear solution techniques. **(iv)** The processes described so far require detailed global information about several parameters, but the availability, acquisition and usage of ancillary data sets about those parameters are additional challenges. **(v)** Because the GNSS-R observations are dependent on the varying instrumental characteristics of the GNSS satellites, uncertainties in determination of these parameters would introduce further errors to the processing of the GNSS-R observations. **(vi)** Moreover, internal instrumental and geometrical parameters of the CYGNSS data products would involve some estimation errors, which should be taken into account. **(vii)** Any GNSS-R based SM retrieval attempt re-

quires comparison to existing SM products as validation reference, but such data sets have their own error budgets and are not the ground truth.

1.3 Overview

This dissertation aims at proposing a learning-based CYGNSS SM retrieval methodology, which is built upon the physical relations between the CYGNSS measurements, SM, and land bio/geophysical parameters, to address as many of the emphasized challenges as possible. The rest of the document describes the development of this method, including the preceding modeling and simulation effort throughout the Ph.D. research as follows:

Chapter 2 describes the bistatic scattering model that has been developed for understanding the physical and electromagnetic mechanisms through not only GNSS-R but also various SoOp configurations. This chapter provides comprehensive theoretical background for modeling of the bistatic scattering as well as vegetation and surface roughness effects.

Chapter 3 gives detailed information about the bistatic simulator framework that is developed based on the bistatic scattering model for the purpose of obtaining a GNSS-R simulation testbed. The chapter also describes how this simulator framework can be exploited for analyzing the GNSS-R observations as well as important signatures under changing values of several configurational, geometrical, and instrumental parameters.

Chapter 4 shares the findings of simulated analyses that were performed by using the bistatic scattering model and its simulator to unveil GNSS-R sensitivity, significant signatures, and dynamic ranges under changing bistatic parameters for determining the requirements and challenges for development of GNSS-R SM retrieval methods.

Chapter 5 provides the details about the development of the physics-aware, learning-based, high spatio-temporal SM retrieval algorithm for CYGNSS observations. It investigates the potential use of several CYGNSS observables in SM retrieval process as well as detailing the challenges for such a task. It explains the selection of the machine learning model and its suitability to solve this non-linear problem. The chapter also gives detailed information about the data acquisition and usage in the algorithm, providing insights into the data distribution statistics.

Chapter 6 concludes the dissertation and its outputs. The conclusion also provides perspectives through potential future work that can be added to the research.

Each chapter from Chapter 2 to Chapter 5 in this dissertation is inherited from the academic articles that have been generated throughout the Ph.D. research, and it has its own comprehensive introduction and literature reviewing materials that are relevant to the particular focus of that chapter. The list of those articles is as follows:

Chapter 2: M. Kurum, M. Deshpande, A. Joseph, P. E. O'Neill, R. H. Lang, and **O. Eroglu**, "SCoBi-Veg: A generalized bistatic scattering model of reflectometry from vegetation for Signals of Opportunity applications," *IEEE Transaction on Geosciences and Remote Sensing*, vol. 57, no. 2, 2018, pp. 1049-1068.

Chapter 3: **O. Eroglu**, D. Boyd, and M. Kurum, "SCoBi: A free, open-source, SoOp Coherent Bistatic Scattering Simulator framework," *IEEE Geoscience and Remote Sensing Magazine*, in press.

Chapter 4: **O. Eroglu**, M. Kurum, and J. Ball, "Response of GNSS-R on dynamic vegetated terrain conditions," *IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing*, vol. 12, no. 5, 2019, pp. 1599-1611.

Chapter 5: **O. Eroglu**, M. Kurum, D. Boyd, and Ali C. Gurbuz, "High spatio-temporal resolution CYGNSS soil moisture estimates using artificial neural networks," *Remote Sensing*, in review.

1.4 Contributions

This dissertation contributes to the field of GNSS-R based SM retrieval by providing a SoOp coherent bistatic scattering model as well as its open-sourced, freely-available, well-documented simulator framework, addressing the requirements and challenges for general and CYGNSS-based SM estimation problems by performing simulated analyses, and proposing a novel physics-aware CYGNSS SM retrieval method that is based on the use of non-parametric, non-linear learning techniques, a few of CYGNSS observables, reference SM data, and a number of ancillary data sets.

Particular contributions can be highlighted as follows:

- Prior to this work, non-parametric, non-linear learning methods have never been employed for estimating SM from CYGNSS observations, although examples of such techniques were used for SM retrieval from observations of other platforms. CYGNSS-based SM retrieval has been commonly attempted by investigating linear regression between CYGNSS observations and reference SM data sets (such as NASA's SMAP SM products) [24, 28, 87].
- To the author's knowledge, this dissertation is the first CYGNSS SM retrieval study to incorporate coherence/incoherence parameters into the retrieval process despite its assumption of the coherent reflections' dominance over land. This approach is inspired by a similar machine learning study that is applied to the CYGNSS-based inundation detection by Rodriguez-Alvarez et al. [135].

- Instead of being only focused on either theoretical modeling, simulation, or SM retrieval from satellite observatory, this study examined a comprehensive research starting from development of a scattering model and finalizing to the development of a SM retrieval algorithm. The physics-aware nature of the proposed learning model has been constructed with the help of the outputs from the simulated analyses that were run through the model and its simulator. Determination of the learning model is also based on the GNSS-R signatures, as well as SM retrieval requirements and challenges, which are observed through the modeling and simulation efforts.

- The existence of the open-source SoOp scattering simulator with its user and developer manuals, and case scenarios, as well as the journal publications about the significant findings throughout the Ph.D. research enables at least partial reproducibility of this dissertation.

CHAPTER 2

SCOBI-VEG: A GENERALIZED BISTATIC SCATTERING MODEL OF REFLECTOMETRY FROM VEGETATION FOR SIGNALS OF OPPORTUNITY APPLICATIONS

This chapter is inherited from the following publication:

M. Kurum, M. Deshpande, A. T. Joseph, P. E. O'Neill, R. H. Lang, and **O. Eroglu**, "SCoBi-Veg: A generalized bistatic scattering model of reflectometry from vegetation for Signals of Opportunity applications," *IEEE Transaction on Geosciences and Remote Sensing*, vol. 57, no. 2, 2018, pp. 1049-1068.

Significant author contributions to the above publication were as follows: Conceptualization and Methodology by Kurum; Investigation and Validation by Kurum and **Eroglu**; Formal analysis by Kurum; Writing by Kurum; Review and Editing by Kurum, **Eroglu**, Deshpande, Joseph, O'Neill, and Lang; Visualization by Kurum; Supervision by Kurum; Project Administration by Kurum.

SCoBi-Veg stands for Signals of opportunity Coherent Bistatic scattering model for Vegetated terrains. It simulates polarimetric reflectometry of vegetation canopy over a flat ground using a Monte Carlo scheme. The model is aimed at assessing the value of navigation and communication satellite Signals of Opportunity in a range of frequencies

from P- to S-bands for remote sensing of a number of geophysical land parameters such as soil moisture and biomass. A fully polarimetric expression for bistatic scattering from a vegetation canopy is first formulated for a general case and is then specialized to the practical case of ground-based/low-altitude platforms with passive receivers overlooking vegetation using the signals transmitted from large distances. Using analytical wave theory in conjunction with distorted Born approximation, the transmit and receive antenna effects (i.e., polarization crosstalk/mismatch, orientation, and altitude) are explicitly accounted for. The forward model developed here enables the understanding of the effect of different geophysical parameters and system configurations on the coherent and incoherent components of the reflected signatures. It can thus help developing robust inverse algorithm for extraction of soil moisture and biomass. The model is applied to P-band signals of geostationary communication satellites to describe polarimetric reflections from tree canopies as observed from down-looking platforms at various altitudes. The relative contributions of diffuse and specular scattering on total reflected power and reflectivity are quantified for various observing scenarios.

2.1 Introduction

Signals of Opportunity (SoOp)-based systems have emerged in recent years as a new domain of microwave remote sensing with great potential to realize the earth science community's need for global geophysical parameter retrieval at high spatiotemporal scales [57, 15]. The key principle of the SoOp approach is to receive and further extract information from free illuminators whose signal reflects off the earth surface. Unlike traditional

microwave remote sensing, the existing signal sources are exploited in bistatic configuration in which the transmitter and the receiver are separated by significant distance. SoOp systems are a powerful, cost-effective approach because they only require the development of a passive receiver, analysis algorithm, and no on-board transmitter. SoOp concept has been widely used for collecting or modeling of Global Navigation Satellite System (GNSS) signal reflections over the ocean surface for estimating wind vectors or altimetry since the 1990s. It has been implemented in technology demonstration experiments, TechDemoSat-1 (TDS-1) [163] and United Kingdom Disaster Monitoring Constellation [61], and a dedicated ocean mission [Cyclone GNSS (CYGNSS)] [140] from space.

The application of GNSS reflectometry (GNSS-R) to other geophysical variables such as soil moisture, vegetation, wetland extent, and snow has also been studied from ground and airborne systems under several different measurement configurations [178]. For instance, interference patterns between direct and reflected signals are exploited using either geodetic or specifically designed GNSS receivers located on a tower to derive soil moisture over bare and vegetated surfaces [134, 4, 39, 97, 25]. In addition, airborne and balloon experiments have been used to evaluate the potential of GNSS-R methodologies for larger scale applications [40, 17]. Most recently, the qualitative analysis of TDS-1 data over land, in conjunction with other satellite products such as soil moisture data from Soil Moisture Ocean Salinity level 3 and normalized difference vegetation index data from Moderate Resolution Imaging Spectroradiometer or Landsat, has demonstrated that the reflected signal has an extremely large dynamic range over land [23, 13, 118]. In addition, the large amount of land observations by the recent CYGNSS mission provides investigators' oppor-

tunities to further infer the sensitivity of GNSS-R measurements to various land features from space.

There is increasing interest to extend these techniques originally developed for L-band GNSS signals to digital transmissions from communication satellites in other bands [151]. For instance, SoOp methods have been applied to communication satellite broadcasts in S- and Ku-bands and used to demonstrate remote sensing of ocean winds, significant wave height, and sea surface height [149, 148]. The P-band signals transmitted by the Military Satellite Communication series of satellites have been proposed to measure root-zone soil moisture through heavy vegetation by recording reflected signals using a simple passive microwave receiver [78, 56, 177]. Furthermore, P-band signals have been used to demonstrate remote sensing of snow water equivalent by analysis of the phase change of the reflected signal [150].

There are a host of navigation and communications satellites illuminating multiple regions of the globe all the time. Furthermore, these satellites span a wide range of the microwave spectrum. Hence, each one of these satellites is a potential candidate for doing different earth science remote sensing. In order to fully leverage the potential of these SoOp transmitters, it is advantageous to develop bistatic scattering models of earth terrain that describe land interactions with the signals at different frequencies in the specular direction. In other words, using scattering models, we can: 1) explore new measurement techniques and configurations; 2) understand the advantages and limitations of each technique; 3) identify the optimum frequencies and signal properties that could help determine the parameters of interest; and 4) deliver sets of test data for training retrieval algorithms.

With these goals in mind, we recently developed a coherent bistatic vegetation scattering model, based on a Monte Carlo simulation, to compute polarimetric bistatic reflectometry [92]. Indeed, due to the scarcity of available bistatic data, it appears that scattering models are a critical step in the advancement of these studies and in the design of future missions or field campaigns.

Despite the somewhat large amount of existing data for soil moisture or vegetation biomass retrieval from monostatic active/passive microwave systems and numerous theoretical models (dating back to the 1970s), bistatic scattering models capable of estimating land parameters from space are still in relatively early stages of development. Very few models which can handle bistatic configurations have so far been developed and can be grouped into two basic categories. The first group [106, 52, 66, 130, 174] is incoherent models that are based on radiative transfer (RT) theory. This approach is a heuristic method based on the law of energy conservation, so it supplies no phase information, but the vector form of RT models is capable of simulating polarimetric data. The second group [92, 155] is coherent models that are based on analytical wave theory in conjunction with the distorted Born approximation and perform coherent sum of the scattered fields. They provide complex field quantities and thus include both amplitude and phase information. This enables one not only to simulate polarimetric bistatic reflectometry for various polarization combinations but also to realize interferometric or beamforming techniques.

In [106], Michigan Microwave Canopy Scattering (MIMICS) model, which is based on the first-order RT theory, is extended to bistatic geometry (hence called bi-MIMICS). The model has been applied to tree canopies at L-, C-, and X-bands with linear polariza-

tions. The bistatic scattering results are compared with those in the backscatter direction to find out optimum sensing configuration for forest biomass estimation. Wu and Jin [174] modified bi-MIMIC model by including combination of circular and linear polarizations of transmiter/receiver to investigate the effect of various observation angles and polarization combinations on the bistatic reflectometry over Aspen stands.

In [52], an RT-based multiple-scattering model (called Tor Vergeta), which is intrinsically bistatic, is extended to consider specular scattering by including circular polarization in specular direction, focusing on GNSS-R applications to biomass monitoring of forests. The model combines the scattering and extinction properties of each scatterer using the matrix doubling algorithm. Guerriero et al. [66] employed the Tor Vergeta model to perform simulations of the scattering coefficient of corn at linear polarizations, over a wide range of observation angles at L- and C-bands. Later, Pierdicca et al. [130] developed a simulator (called SAVERS) for GNSS reflections from bare and vegetated soils, which uses the Tor Vergeta model with circular polarization. The polarization synthesis technique, applied to take antenna polarization mismatch and crosstalk into account, has been presented. Preliminary assessment of the simulator is successfully demonstrated against the field data acquired over sunflower plants during a growing cycle.

Thirion-Lefevre et al. [155] studied bistatic scattering by forested areas using a coherent scattering model called COBISMO, which is an extension of previous developed coherent backscatter model. They considered linear polarization and analyzed radiometric and polarimetric aspects of bistatic scattering coefficients of forest canopies at P-band to check the physical relevance of the model behavior. The present model, Signals of oppor-

tunity Coherent Bistatic scattering model for Vegetated terrains (SCoBi-Veg), falls in the same category with COBISMO, but includes more comprehensive antenna characteristics as well as circular polarization since the SoOp transmitters are often circularly polarized and the receiver antennas cannot be constructed to produce pure polarization states.

The SCoBi-Veg model calculates the received complex field in three main contributions: 1) direct term; 2) specular term; and 3) diffuse term by explicitly accounting for both antenna and scene characteristics. The first contribution represents the line of sight or the shortest path between the antennas while the second term denotes the scattering along the specular direction. Finally, the diffusely scattered waves arrive at the receiver antenna from a wide range of angles in both azimuth and elevation due to scattering from the illuminated volume. An average diffuse term is obtained by a sufficient number of realizations of vegetation through Monte Carlo simulations. In received reflected field, the antenna characteristics and orientation play a key role. For instance, for ground-based systems, the antenna radiation pattern projected on the surface is not uniformly distributed in phase, amplitude, or polarization.The bistatic radar coefficient calculated by the model thus needs to have the antenna characteristics embedded in it as well as the statistical and physical properties of the terrain to mimic the real measurement setting. Previous studies have usually assumed plane wave illumination/scattering, and/or ignored such antenna effects. The SCoBi-Veg model considers variations of both the strength and polarization states of the received wave along the beam direction so that the same model could be uniformly applied across different platforms at various altitudes. This is an important step

since the groundbased receivers are often used for validation of algorithms to be used by spaceborne observations.

This paper aims at presenting an overview of the newly formulated coherent bistatic vegetation scattering model, i.e., SCoBi-Veg, as well as P-band analysis of polarimetric specular and diffuse contributions for a tree canopy. A companion paper is currently in preparation to provide details of open-source implementation of the SCoBi-Veg model in MATLAB/Octave environment. Section 2.2 starts providing explicit expressions of direct, specular, and diffuse contributions to the received field in a general bistatic configuration. The field quantities are then converted to the polarimetric received power in modified Stokes vectors. Section 2.2 concludes by specializing the model for geometries where the relative distance of the transmitter to the receiver with respect to the specular ground point is large. This represents the most common observation configuration for SoOp concepts. Section 2.3 first introduces the simulation setting and parameters at P-band, and then provides simulation results and discussion of the results. Finally, Section 2.4 summarizes the important aspects of the model and draws conclusions based on the simulated results. Four appendixes are provided at the end to supplement understanding of the model formulation, definitions, and notations.

2.2 Model

In this section, the details of the formulation behind the SCoBi-Veg model will be presented. A general bistatic scattering case is first considered, where two antennas do not have their main beam axis pointing at each other and both of which are overlooking

the vegetation. One antenna system is associated with the SoOp transmitter while the other refers to the passive receiver. The vegetation is represented as ensemble of canonical scatterers located above a flat ground. The model calculates direct, specular, and diffuse components of the received complex field for a wide-beam antenna system with nonzero cross-polarization. The model is specialized to the most common SoOp application, where the transmitter is located far away. The local incident angles are assumed to be constant (parallel incident rays) and spreading loss effects due to the incoming wave are ignored, but the spreading loss and sphericity of the scattered wave are considered due to proximity of the receiver platforms that may operate close to the ground. This configuration will represent plane wave incidence and spherical wave scattering. Furthermore, the formulation considers variations of both the strength and polarization states of the received wave along the beam direction by taking into account for the polarization mismatch and crosstalk. Since the model preserves the phase of the scattered field from the canopy, it allows us to calculate various combination of transmitreceive polarization combinations to investigate various polarization signatures.

A. General Case

In this section, we will consider a general bistatic scattering case where two antennas do not have their main beam axis pointing at each other. A step-by-step process will be outlined to obtain the received field when two nonideal antennas are mismatched in polarization. A transmit antenna located at point T (at a height of h_t) is considered to be illuminating the earth surface and another antenna as a receiver located at point R (at a height of h_r) is collecting the scattered field from the surface as depicted in Figure 2.1.

In this paper, the earth surface is assumed to be planar, mostly valid for ground-based or low altitude aircraft platforms, while the same formulation can be extended for spaceborne geometries, where the earth curvature is important. In the reference coordinate system (*x, y, z*), the zenith (or surface normal $\hat{\mathbf{n}}$) is in $\hat{\mathbf{z}}$ direction while the unit vectors $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ form the ground (*xy* plane) and are aligned with the local East and North directions, respectively. The reference frame defined here is similar to the standard local east, north, up (ENU) system. The transmit antenna in a coordinate system (x_t, y_t, z_t) points to the ground (point B_t) with the incidence angle $\theta_{0t} = cos^{-1}(-\hat{n} \cdot \hat{z}_t)$ while the receive antenna in a coordinate system (x_r, y_r, z_r) points to the ground (point B_r) with the incidence angle $\theta_{0r} = cos^{-1}(\hat{n} \cdot \hat{z}_r)$. The main beam (boresight) axis of both antennas is aligned with their own local *z* coordinates. For notational clarity, unit vectors are bolded with a hat over them while vectors with magnitudes other than one will be bolded only. In addition, the subscript *r* refers to a quantity associated with the receive antenna, and *t* is associated with the transmit antenna.

The electric field radiated by an antenna is commonly defined by a spherical coordinate system. The incident wave, having θ_t and φ_t components in transmit antenna frame, travels in the radial direction $\hat{k}_i$ from the transmitter, while the scattered wave, having θ_r and φ_r components in receive antenna frame, travels in the radial direction $\hat{k}_o$ toward the receiver. One can write

$$\hat{k}_i = \hat{x}_t sin\theta_t cos\varphi_t + \hat{y}_t sin\theta_t sin\varphi_t + \hat{z}_t cos\theta_t \tag{2.1a}$$

$$-\hat{k}_o = \hat{x}_r sin\theta_r cos\varphi_r + \hat{y}_r sin\theta_r sin\varphi_r + \hat{z}_r cos\theta_r \tag{2.1b}$$

where the subscript *i* refers to a quantity associated with incident (transmitted) wave, and *o* is associated with the scattered (received) wave. The orthonormal unit vectors $(\hat{x}_t, \hat{y}_t, \hat{z}_t)$ and $(\hat{x}_r, \hat{y}_r, \hat{z}_r)$ are defined along the antenna frame axes.

The point *S* in Figure 2.1 denotes the location of the point of reflection that follows the shortest multipath distance from ground. This is called the "Specular Point," where the incident and reflected waves make the same angle of θ_s with the surface normal. It is determined from simple image theory by confining the discussion to a plane earth. A local coordinate system (x', y', z') centered at the specular point *S* is also shown. The x'-axis lies along the ground range between the source and the receiver (between projected points G_t and G_r) while the z'-axis is normal to the surface. The angle between the local x'-axis and the reference (ENU) x-axis, $\varphi = cos^1(\hat{x} \cdot \hat{x}')$, defines the azimuth direction of the transmit antenna from the local East. The orthonormal unit vectors $(\hat{x}'), \hat{y}'), \hat{z}'))$ are defined along the local coordinate axes.

In reality, the energy source illuminates a large region of the surface, depending on the extent of the transmit antenna main beam. Radiation traveling from surface points by any other path than the specular point will travel a longer distance. The difference between any other path length and the shortest distance will form a family of ellipses on the ground. The elliptic zones can be defined around the specular point with unique phase differences (multiples of half-wavelengths) called Fresnel zones. The exact dimensions of the Fresnel ellipses are given in [83] for planar earth. Depending on the surface roughness and veg-

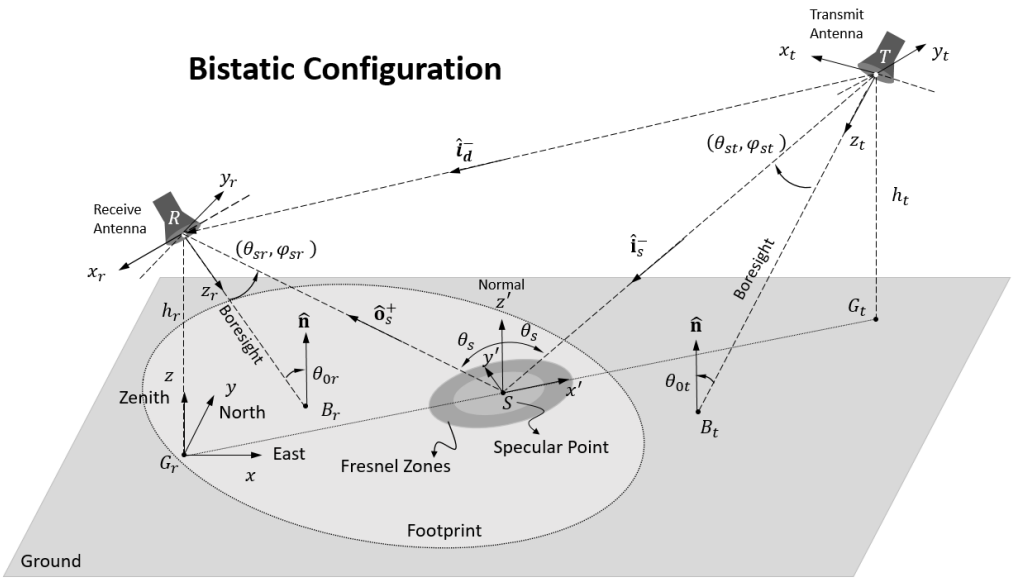

Figure 2.1: Arbitrary bistatic antenna configuration.

etation, the number of contributing zones may differ. In addition, the signal received by any SoOp instrument will be in general constrained by footprints of both transmitter and receiver range and Doppler discriminations.

In the present investigation, P-band geostationary transmitter and tower-based receiver are considered as an example scenario (without loss of generality), so both platforms are not moving. Thus, there will be no Doppler shift in the received field. Since the model preserves the phase, Doppler effects can be incorporated in the simulations as well for GNSS-R applications. Thus, only footprint of receiver and range discrimination are considered in the simulations. Range discrimination represents the relative delay with respect to the specular reflection point. As a result, Fresnel zones correspond to range delays that the SoOp instrument discriminates around the specular point. Fresnel zone basically refers

to iso-range discrimination areas proportional to a fraction of the wavelength, which can also be based on the chip resolution as done in the GNSS-R community.

Let us now situate some vegetation over the ground plane. In Figure 2.2, bistatic scattering over a vegetated landscape is illustrated from a perspective (looking along $\hat{y}'$) that is perpendicular to the plane containing the receiver, transmitter, and specular point. The canopy is represented as an ensemble of canonical scatterers located above the ground plane. This approach is called discrete scatterer where the canopy can consist of randomly distributed branches, stems, leaves, needles, and vertically distributed tree trunks or stalks all having prescribed location and orientation statistics. Leaves are modeled as flat thin dielectric discs [104, 105] and branches, trunks, and stalks are modeled as finitely long lossy-dielectric cylinders [147, 79]. The single-scattering characteristics of these constituents when averaged determine the attenuation and scattering properties of the canopy. The advantage of the discrete approach is that the results are expressed in terms of quantities (plant geometry and orientation statistics) that are related to the biophysical properties of individual plants. These anonical scatterers are assumed to be uniformly distributed throughout the layer.

For sake of convenience, the local coordinate system (x', y', z') is moved to the top surface of the vegetation. The ground (z = 0) in ENU is thus denoted by $z' = d$ plane in local coordinates. The transmit and receive antennas are located at x_i and x_o, respectively. The total received field will have direct and multipath components

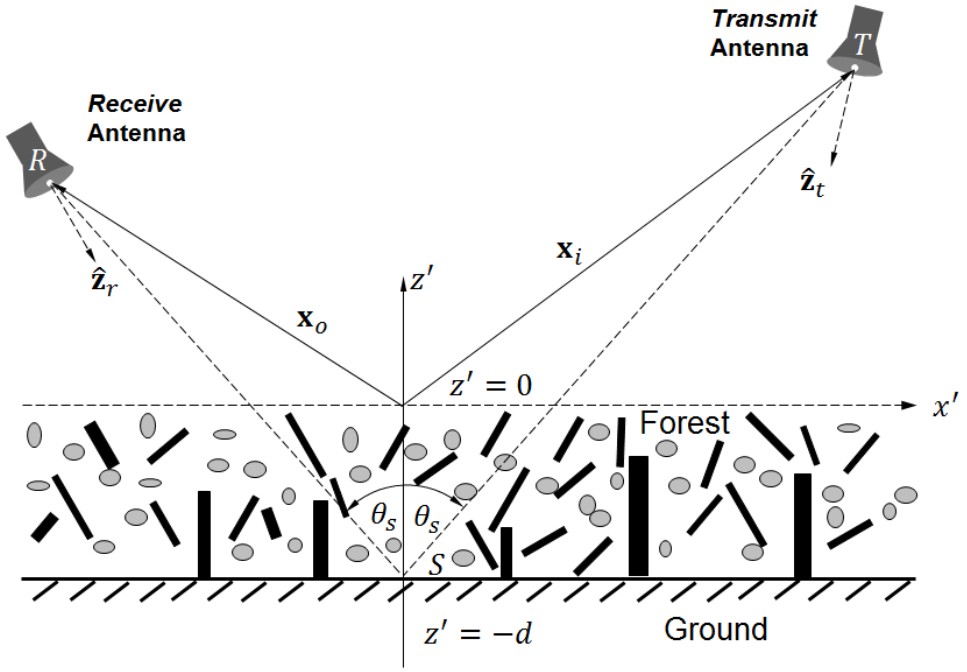

Figure 2.2: Representation of a vegetated landscape.

$$\underline{b} = \begin{bmatrix} b_1 \\ \\ b_2 \end{bmatrix} = \underline{b}_d + \underline{b}_m \tag{2.2a}$$

where the direct component is

$$\underline{b}_d = \begin{bmatrix} b_{1d} \\ \\ b_{2d} \end{bmatrix} \tag{2.2b}$$

and the multipath component is

$$\underline{b}_m = \begin{bmatrix} b_{1m} \\ \\ b_{2m} \end{bmatrix} \tag{2.2c}$$

The elements of $\underline{b}$ vector (i.e., b_1 and b_2) represent network scattering parameters at the physical antenna ports such as port 1 (along x_r-axis) and port 2 (along y_r-axis),

respectively, when the antenna is in reception mode. They are complex voltage quantities and their magnitude squares are equal to received power at each port [171]. Throughout the text, the lowercase letters with a single underline denote two-by-one vectors while two-by-two matrices are likewise indicated using lowercase letters with a double underline.

There are two different uses of Coherent in this paper. First, with the coherent model, it is meant that the absolute phase (complex electric field) of the received signal is preserved. This means that the scattering considers relative positions of scatterers as well as phase change due to reflection and scattering since they are added coherently (summing in complex electric field). Doppler is another source of phase shift, but not considered in the present investigation. Second, with the coherent term as described below, the specular reflection contribution is referred, which comes from a single point only from geometric optics point of view. In other words, the coherent term is simply the expected value of the field.

The multipath component will have coherent (specular reflection) and incoherent (off-specular or diffuse scattering) contributions

$$\underline{b}_m = \underline{b}_m^{coh} + \underline{b}_m^{inc} \tag{2.3a}$$

where the incoherent contribution is

$$\underline{b}_m^{inc} = \begin{bmatrix} b_{1m}^{inc} \\ b_{2m}^{inc} \end{bmatrix} \tag{2.3b}$$

and the coherent contribution is

$$\underline{b}_m^{coh} = \begin{bmatrix} b_{1m}^{coh} \\ \\ b_{2m}^{coh} \end{bmatrix} \tag{2.3c}$$

Below, each contribution will be described separately.

1) Direct Contribution: This contribution represents the shortest path where the transmitted signal arrives at the receiving antenna directly and is given by

$$\underline{b}_d = K\frac{e^{ik_0 r_d}}{r_d}\underline{\underline{g}}_r(\hat{k}_d)\cdot\underline{\underline{u}}_{t\to r}(\hat{k}_d)\cdot\underline{\underline{g}}_t(\hat{k}_d)\cdot\underline{e}_t \tag{2.4a}$$

where $\hat{k}_d$ is the unit vector from the transmit antenna to the receiver antenna and the constant *K* is given by

$$K = i\frac{\lambda_0}{4\pi}\sqrt{G_{0r}EIRP} \tag{2.4b}$$

where $i = \sqrt{-1}$ is the unit complex number, EIRP = $P_t G_{0t}$ is the equivalent isotropic radiated power (EIRP), λ_0 is the wavelength in free space, P_t is the power transmitted, G_{0t} is the maximum gain of the transmit antenna, G_{0r} is the maximum gain of the receive antenna, r_d is the slant rangethe distance between the transmit and receive antennas, and k_0 is the free-space wave number, i.e. $k_0 = 2/\lambda_0$.

It is important to realize that antennas cannot be constructed to produce pure polarization states [116]. There will always be some nonzero cross-polarization level (crosstalk isolation). Nevertheless, we refer to the antenna, for example, as linearly polarized. Here, it is recognized that we mean nominally linear (horizontal) and a cross-polarized (vertical)

components will be present. Similarly, for a circularly polarized antenna, we mean nominally right-hand circularly polarized (RHCP) and a cross-polarized [left-hand circularly polarized (LHCP)] components will be present. As discussed in Appendix A, the pattern matrices can be constructed to take into account the cross-polarization leakage between ports of the antennas. In (2.4a), the receive and transmit antenna patterns are represented by the two-by-two matrices g_r and g_t , called normalized voltage pattern matrices, respectively. Their expressions are given in (A.3) and (A.6) for linearly and circularly polarized antennas, respectively.

The vector $\underline{e}_t = [e1\, e2]^T$ in (2.4a), where the superscript T is transpose, defines the nominal polarization state of the transmit antenna ports, and its magnitude is equal to 1. It represents the polarization state in the main beam direction of the transmit antenna. The polarization, however, varies for increased off-axis angles over the radiation pattern. Due to the orientation of the transmit and the receive antennas, the antenna polarization states are different off the boresight and this will introduce polarization mismatch (hence loss). One can relate polarization states of the transmit and receive antennas through a polarization (field) rotation matrix (see Appendix B). For instance, the rotation matrix from circularly polarized transmit antenna basis ($\hat{u}_{t1}^R$, $\hat{u}_{t2}^L$) to linearly polarized receive antenna basis ($\hat{u}_{r1}^X$, $\hat{u}_{r2}^Y$) can be written as

$$\underline{\underline{u}}_{t\to r}(\hat{k}_d) = \begin{bmatrix} \hat{u}_{t1}^R \cdot \hat{u}_{r1}^{X*} & \hat{u}_{t2}^L \cdot \hat{u}_{r1}^{X*} \\ \hat{u}_{t1}^R \cdot \hat{u}_{r2}^{Y*} & \hat{u}_{t2}^L \cdot \hat{u}_{r2}^{Y*} \end{bmatrix} \tag{2.4c}$$

where the symbol $*$ represents complex conjugate, the superscripts R, L, X, and Y denote RHCP, LHCP, linear polarization in X-direction (along x-axis port), and linear polarization in Y -direction (along y-axis port), respectively, as defined in A. In this paper, the antenna y-axis of both transmit and receive frames is chosen to always be parallel to the ground, so that it represents the horizontal polarized port while the other port (x-axis) represents vertical polarization when the antenna is linearly polarized. This scheme is accomplished through antenna rotation matrices as described in C.

1) Specular Reflection (Coherent) Contribution: The coherent contribution represents the reflection from the specular point S on the ground. According to image theory under the assumption that the surface is of infinite extent, it can be written as

$$\underline{b}_m^{coh} = K \frac{e^{ik_0(r_{st}+r_{sr})}}{(r_{st}+r_{sr})} \underline{\underline{g}}_r(\hat{o}_s^+) \cdot \underline{\underline{u}}_{s\to r}(\hat{o}_s^+) \cdot \underline{\underline{r}}_s(\hat{o}_s^+, \hat{i}_s^-) \cdot \underline{\underline{u}}_{t\to s}(\hat{i}_s^-) \cdot \underline{\underline{g}}_t(\hat{i}_s^-) \cdot \underline{e}_t \tag{2.5}$$

where the constant K is defined above in (2.4b), and the distances r_{st} and r_{sr} are the ranges from the transmit antenna to the specular point and from the receive antenna to the specular point, respectively.

The wave attenuates as it propagates down from the top of the vegetation to the ground. It then reflects from the surface specularly, and it again attenuates from the ground to the top of the vegetation. The coherent term is essentially attenuated twice by the vegetation. The attenuation and phase change of the coherent wave, propagating in the equivalent medium, is found by calculating the mean field within the medium [96, 95] according to FoldyLax approximation [53, 102]. This approach assumes single scattering and azimuthal

uniformity, but inhomogeneous along vertical axes. The vegetation canopy can be divided into stratified layers to account for variability along vertical axes. The paulownia trees used in this paper are divided into four layers (one trunk and three canopy layers) as described in Section 2.3.

The propagation and reflection process within the vegetation can be written as a specular reflection matrix

$$\underline{\underline{r}}_s(\hat{o}_s^+, \hat{i}_s^-) = \underline{\underline{t}}(\hat{o}_s^+) \cdot \underline{\underline{r}}_g(\theta_s) \cdot \underline{\underline{t}}(\hat{i}_s^-) \tag{2.6}$$

where $\underline{\underline{t}}$ is the vegetation transmission matrix and $\underline{\underline{r}}_g$ is the ground reflection matrix. The unit vectors $\hat{i}_s^-$ and $\hat{o}_s^+$ describe the wave propagation in the incoming and outgoing directions, respectively. The superscripts $+$ and $-$ are used to denote the wave modes propagating in the positive and negative z-directions. The subscript s indicates that quantities are calculated for the specular point. The angle θ_s is the angle of reflection at the specular point and can be calculated as

$$\theta_s = cos^{-1}\{-\hat{n} \cdot \hat{i}_s^-\} = cos^{-1}\{\hat{n} \cdot \hat{o}_s^+\} \tag{2.7}$$

The transmission matrix is responsible for the incident and scattered waves to experience attenuation when they travel in the mean medium [95]. Under the assumption that the scatterers are uniformly distributed in azimuth, the following expression for the one-way transmission matrix is obtained:

$$\underline{\underline{t}}(\hat{i}) = \begin{bmatrix} e^{i\Delta\kappa_{zp}(\hat{i})d} & 0 \\ 0 & e^{i\Delta\kappa_{zq}(\hat{i})d} \end{bmatrix} \tag{2.8a}$$

where d is the depth of the vegetation, $\hat{i}$ is a unit vector in the direction of propagation (i.e., $\hat{i}_s^-$ or $\hat{o}_s^+$), and the exponents of matrix elements are of the form (FoldyLax approximation) [53, 102]

$$\Delta\kappa_{zq}(\hat{i}) = \frac{2\pi}{k_0 cos\theta} \sum_{\alpha} \rho_\alpha \bar{f}_{qq}^{\alpha}(\hat{i}; \hat{i}; \beta_\alpha) \tag{2.8b}$$

where the angle θ is the angle between the unit vector $\hat{i}$ and z-axis. The summation index α denotes the scatterer types such as leaves, needles, branches, stalks, or trunks. The number density of the scatterer type α is given by ρ_α. Here, the quantity $\bar{f}_{pq}^{\alpha}(\hat{o}; \hat{i}; \beta_\alpha)$ denotes the bistatic scattering amplitude of the scatterer type α, where $\hat{i}$ is in the direction of the incident wave and $\hat{o}$ is in the direction of the observation point. The subscripts p and q can be horizontal (H) or vertical (V), and as a result, co- and cross-polarized cases are treated simultaneously.

The quantity β_α describes the orientation, size, and position statistics of the scatterer type. The bar over the scattering amplitude denotes ensemble average over the angular and size statistics of the particles. In this paper, the leaves (or needles) are represented by an average-size circular disk (or circular finite cylinder); hence, the averaging is done for orientation angles only. The trunks (or stalks) are vertical and have a typical size. No averaging is, therefore, performed on trunks (or stalks). The branches (if different kinds are

included) can be divided into several groups, and each group can have an average length and an average diameter. Orientation averaging can then be performed on each branch group. In other words, orientation of leaves and branches is random but dimension is fixed, trunks dimensions and orientation are fixed, and position of each scatterer is random. Notice that in (2.8b), $\bar{f}_{qq}^{\alpha}(\hat{i};\hat{i};\beta_{\alpha})$ is the forward scattering amplitude of the α^{th} group of scatterers.

In (2.6), the ground reflection matrix is given by

$$\underline{\underline{r}}_g(\theta_S) = \begin{bmatrix} \Gamma_{gp}(\theta_s) & 0 \\ 0 & \Gamma_{gq}(\theta_s) \end{bmatrix} \tag{2.9a}$$

It is assumed that the rough surface under the vegetation is smooth and follows Kirchhoff's approximation with a Gaussian height distribution [26]; therefore, diffuse scattering from ground is not accounted for and the reection coefficient of the rough surface is expressed with the Fresnel coefficients as

$$\Gamma_{gq}(\theta_S) = r_{gq} e^{-2(k_0 s\cos\theta_s)^2} \tag{2.9b}$$

where Γ_{gq} is the q-polarized Fresnel reflection coefficient of the average dielectric surface, s is the surface rms height, and θ_S is the angle of reflection from the surface, defined in (2.7). The diffuse surface scattering is attenuated by vegetation twice, and it is not expected to be a strong contributor to the total power unless it is very rough surface and topographic relief exists.

The polarization basis of the receive antenna and the specular ground point along scattered wave ($\hat{o}_s^+$), denoted by the subscript $s \rightarrow r$, are related by a rotation matrix (see Appendix B) that is given by

$$\underline{\underline{u}}_{s\rightarrow r}\left(\hat{\mathbf{o}}_s^+\right) = \begin{bmatrix} \hat{\mathbf{v}}_o^s \cdot \hat{\mathbf{u}}_{r1}^* & \hat{\mathbf{h}}_o^s \cdot \hat{\mathbf{u}}_{r1}^* \\ \hat{\mathbf{v}}_o^s \cdot \hat{\mathbf{u}}_{r2}^* & \hat{\mathbf{h}}_o^s \cdot \hat{\mathbf{u}}_{r2}^* \end{bmatrix} \tag{2.10a}$$

where the local horizontal and vertical polarization in the direction from the specular point to the receiver are, respectively, given by

$$\hat{\mathbf{h}}_o^s = \frac{\hat{\mathbf{n}} \times \hat{\mathbf{o}}_s^+}{|\hat{\mathbf{n}} \times \hat{\mathbf{o}}_s^+|} \text{ and } \hat{\mathbf{v}}_o^s = \hat{\mathbf{h}}_o^s \times \hat{\mathbf{o}}_s^+ \tag{2.10b}$$

where $\hat{\mathbf{n}} = \hat{z}'$ and the horizontal polarization vector $\hat{\mathbf{h}}_o^s$ is parallel to the local $x'y'$ plane.

The polarization basis of the transmit antenna and the specular point along the incident wave ($\hat{i}_s^-$), denoted by the subscript $t \rightarrow s$, are related by a rotation matrix (see Appendix B) that is given by

$$\underline{\underline{u}}_{t\rightarrow s}\left(\hat{\mathbf{i}}_s^-\right) = \begin{bmatrix} \hat{\mathbf{u}}_{t1} \cdot \hat{\mathbf{v}}_i^{s*} & \hat{\mathbf{u}}_{t2} \cdot \hat{\mathbf{v}}_i^{s*} \\ \hat{\mathbf{u}}_{t1} \cdot \hat{\mathbf{h}}_i^{s*} & \hat{\mathbf{u}}_{t2} \cdot \hat{\mathbf{h}}_i^{s*} \end{bmatrix} \tag{2.11a}$$

where the local horizontal and vertical polarization in the direction from transmitter to the specular point are, respectively, given by

$$\hat{\mathbf{h}}_i^s = \frac{\hat{\mathbf{i}}_s^- \times \hat{\mathbf{n}}}{\left|\hat{\mathbf{i}}_s^- \times \hat{\mathbf{n}}\right|} \text{ and } \hat{\mathbf{v}}_i^s = \hat{\mathbf{i}}_s^- \times \hat{\mathbf{h}}_i^s \tag{2.11b}$$

where the horizontal polarization vector $\hat{\mathbf{h}}_i^s$ is parallel to local $x'y'$ plane.

3) Diffuse Scattering (Incoherent) Contribution: The diffusely scattered waves arrive at the receiver antenna from a wide range of angles in both azimuth and elevation due to scattering from the illuminated volume within the contributing Fresnel zones. In "discrete scatterer" modeling [95, 156], the vegetation layer is usually replaced by a slab of equivalent homogenous medium by the FoldyLax theory. Using analytical wave theory in conjunction with distorted Born approximation [95], the transmit and receive antenna effects (i.e., polarization, orientation, and shape) can be explicitly accounted for.

Let us consider the geometry of the problem, shown in Figure 2.3, where a single scatterer of type α is embedded at point P_α in the medium over the half-space. The position vector is given by $r_\alpha = x_\alpha\hat{\mathbf{x}}'+y_\alpha\hat{\mathbf{y}}'z_\alpha\hat{\mathbf{z}}'$ in local coordinates. An expression for the received field due to this single particle was obtained by employing distorted Born approximation in [91] with the assumption that the regular far-field conditions hold, i.e., each scatterer in the vegetation layer is in the far-zone of the transmit antenna and the receive antenna is also in the far-zone of each scatterer within the layer. The energy source illuminates a large region of surface, inducing currents that radiate in all directions. At any point in space, the reflected field is sum of the radiation from these induced currents according to the Huygens principle. The scattered plane wave of propagation arises as a consequence of all incident plane waves of differing propagation vectors being scattered by the medium into observation direction. If the scatterer is in the far field of the transmit antenna, the double integral involving the transverse wavenumber of the incident wave can be evaluated asymptotically by the method of stationary phase. In addition, with an assumption on the particles such

that each particle is uniformly illuminated by the incident wave, the magnitude of the mean wave is essentially constant in the vicinity of the particle. In addition, the antennas with narrow beam and pure polarization state were considered in [91].

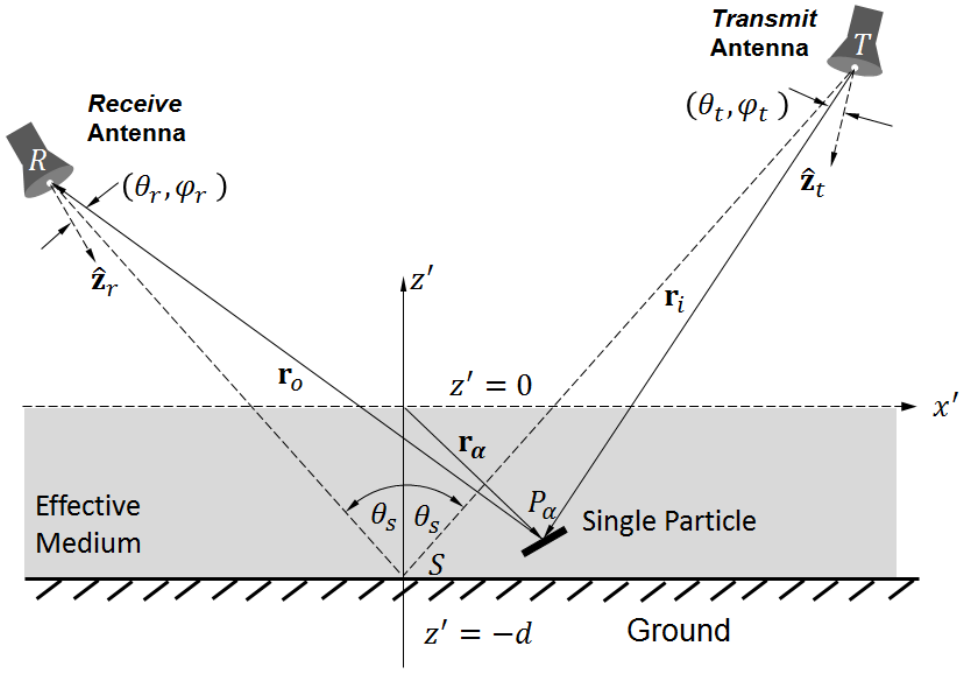

Figure 2.3: Scattering from a single particle immersed in the slab of mean medium over a flat ground surface.

The results in [91] are here adapted to a general case where both antennas are considered to be wide-beam with nonzero cross-polarized components. This implies that the strength and polarization of the transmitted and received waves depend on the antenna properties along the beam direction. Summing the result of the single particle scattering over all types and particles and taking into account the polarization mismatch (rota-

tion matrices) and the polarization crosstalk (antenna pattern matrices with co- and cross-polarization components) lead to the following:

$$\begin{aligned}\underline{b}_m^{\text{inc}} =& K\frac{e^{ik_0(r_{st}+r_{sr})}}{r_{st}r_{sr}}\sum_{\alpha}\sum_{n=1}^{N_a}\sum_{x}B_{\alpha,n}^{x} \\ &\times\left[\underline{\underline{g}}_r\left(\hat{\mathbf{o}}_x\right)\cdot\underline{\underline{u}}_{p\to r}\left(\hat{\mathbf{o}}_x\right)\cdot\underline{\underline{s}}_{\alpha,n}^{x}\left(\hat{\mathbf{o}}_x,\hat{\mathbf{i}}_x\right)\cdot\underline{\underline{u}}_{t\to p}\left(\hat{\mathbf{i}}_x\right)\cdot\underline{\underline{g}}_t\left(\hat{\mathbf{i}}_x\right)\cdot\underline{e}_t\right]\end{aligned} \tag{2.12}$$

where the constant K and the distances r_{st} and r_{sr} are defined above. The summation indices α, n, and x denote the types of particles, particle number of type α, and the scattering mechanisms, respectively. N_α is the total number of particles of type α within the illuminated volume around the specular point. The normalized "voltage" pattern matrices, $\underline{\underline{g}}_t$ and $\underline{\underline{g}}_r$, of transmit and receive antennas are given in Appendix A and can be linearly or circularly polarized as discussed earlier. The strengths of the incident and scattered waves at each particle are directly proportional to the value of the antenna patterns along that direction.

The wave from each particle follows four distinct paths before arriving at the receiver, denoted by xdd, rd, dr, rr. Each mechanism is described by the following scattering matrices:

$$\underline{\underline{s}}_{\alpha,n}^{\text{dd}}\left(\hat{\mathbf{o}}_\alpha^+,\hat{\mathbf{i}}_\alpha^-\right)=\underline{\underline{t}}\left(\hat{\mathbf{o}}_\alpha^+\right)\cdot\underline{\underline{f}}_{\alpha}\left(\hat{\mathbf{o}}_\alpha^+,\hat{\mathbf{i}}_\alpha^-;\beta_{\alpha,n}\right)\cdot\underline{\underline{t}}\left(\hat{\mathbf{i}}_\alpha^-\right) \tag{2.13a}$$

$$\underline{\underline{s}}_{\alpha,n}^{\text{dr}}\left(\hat{\mathbf{o}}_\alpha^+,\hat{\mathbf{i}}_{\alpha I}^+\right)=\underline{\underline{t}}\left(\hat{\mathbf{0}}_\alpha^+\right)\cdot\underline{\underline{f}}_{\alpha}\left(\hat{\mathbf{o}}_\alpha^+,\hat{\mathbf{i}}_{\alpha I}^+;\beta_{\alpha,n}\right)\cdot\underline{\underline{r}}_{g}\left(\hat{\mathbf{i}}_{\alpha I}^+\right)\cdot\underline{\underline{t}}\left(\hat{\mathbf{i}}_{\alpha I}^+\right) \tag{2.13b}$$

$$\underline{\underline{s}}^{\mathrm{rd}}_{\alpha,n}\left(\hat{\mathbf{o}}^{-}_{\alpha I},\hat{\mathbf{i}}^{-}_{\alpha}\right)=\underline{\underline{t}}\left(\hat{\mathbf{o}}^{-}_{\alpha I}\right)\cdot\underline{\underline{r}}_{g}\left(\hat{\mathbf{o}}^{-}_{\alpha I}\right)\dot{=}\underline{\underline{f}}_{\alpha}\left(\hat{\mathbf{o}}^{-}_{\alpha I},\hat{\mathbf{i}}^{-}_{\alpha};\beta_{\alpha,n}\right)\cdot\underline{\underline{t}}\left(\hat{\mathbf{i}}^{-}_{\alpha}\right) \tag{2.13c}$$

$$\underline{\underline{s}}^{\mathrm{rr}}_{\alpha,n}\left(\hat{\mathbf{o}}^{-}_{\alpha I},\hat{\mathbf{i}}^{+}_{\alpha I}\right)=\underline{\underline{t}}\left(\hat{\mathbf{o}}^{-}_{\alpha I}\right)\cdot\underline{\underline{r}}_{g}\left(\hat{\mathbf{o}}^{-}_{\alpha I}\right)\cdot\underline{\underline{f}}_{\alpha}\left(\hat{\mathbf{o}}^{-}_{\alpha I},\hat{\mathbf{i}}^{+}_{\alpha I};\beta_{\alpha,n}\right)\dot{\underline{\underline{r}}}\,g\left(\hat{\mathbf{i}}^{+}_{\alpha I}\right)\cdot\underline{\underline{t}}\left(\hat{\mathbf{i}}^{+}_{\alpha I}\right) \tag{2.13d}$$

where the unit vectors are defined as

$$\hat{\mathbf{i}}^{-}_{\alpha}=\frac{\mathbf{r}_i}{r_i},\quad r_i=|\mathbf{r}_i|=|\mathbf{r}_\alpha-\mathbf{x}_i| \tag{2.14a}$$

$$\hat{\mathbf{o}}^{+}_{\alpha}=\frac{\mathbf{r}_o}{r_o},\quad r_o=|\mathbf{r}_o|=|\mathbf{r}_\alpha-\mathbf{x}_o| \tag{2.14b}$$

$$\hat{\mathbf{i}}^{+}_{\alpha I}=\frac{\mathbf{r}_{iI}}{r_{iI}},\quad r_{iI}=|\mathbf{r}_{iI}|=|\mathbf{r}_\alpha-\mathbf{x}_{iI}| \tag{2.14c}$$

$$\hat{\mathbf{o}}^{-}_{\alpha I}=\frac{\mathbf{r}_{oI}}{r_{oI}},\quad r_{oI}=|\mathbf{r}_{oI}|=|\mathbf{r}_a-\mathbf{x}_{oI}| \tag{2.14d}$$

and where the superscripts $+$ and $-$ are used to denote the wave modes propagating in the positive and negative z-directions, respectively. The subscript I is introduced in the notation to refer to the auxiliary image antenna.

Figure 2.4 illustrates the scattering mechanisms for a single particle. The reflected paths are illustrated by the image antenna representation, where the images of the transmit

and receive antennas with respect to ground are positioned at $\mathbf{x}_{iI} = \mathbf{x}_i - 2(\mathbf{x}_i \cdot \hat{\mathbf{z}}' + d)\hat{\mathbf{z}}'$ and $\mathbf{x}_{oI} = \mathbf{x}_o - 2(\mathbf{x}_o \cdot \hat{\mathbf{z}}' + d)\hat{\mathbf{z}}'$, respectively. In order to account for the attenuation of the reflected components, an image medium is included in Figure 2.4. Each scattering mechanism can be visualized as follows.

1) Single Bounce Contribution [see (2.13a)*]*: Directdirect (dd) term follows the path $TP_\alpha R$. The field travels in the direction of $\hat{\mathbf{i}}_x = \hat{\mathbf{i}}_\alpha$ toward the particle of type α, located at r_α, and then the incident field is bistatically scattered from this particle and travels in the direction of $\hat{\mathbf{o}}_x = \hat{\mathbf{o}}_\alpha^+$ toward the receive antenna.

2) Double Bounce Contributions [see (2.13b) *and* (2.13c)*]*: It includes two mechanisms such as reflected-direct (dr) and direct-reflected (rd) terms. The first mechanism follows the path $T_I P - \alpha R$. The field incident on the particle in the direction of $\hat{\mathbf{i}}_x = \hat{\mathbf{i}}_{\alpha I}^+$ from the image of the transmit antenna is reflected from the ground and bistatically scattered from the particle in the direction of $\hat{\mathbf{o}}_x = \hat{\mathbf{o}}_\alpha^+$ toward the receive antenna. The second mechanism follows the path $TP_\alpha R_I$. The field incident on the particle in the direction of $\hat{\mathbf{i}}_x = \hat{\mathbf{i}}_\alpha$ is bistatically scattered from the particle in the direction of $\hat{\mathbf{o}}_x = \hat{\mathbf{o}}_{\alpha I}^-$ toward the image of receive antenna after reflection from the ground.

3) Triple Bounce Contribution [see (2.13d)*]*: Reflected-reflected (rr) term follows the path $T_I P_\alpha R_I$. The field incident on the particle in the direction of $\hat{\mathbf{i}}_x = \hat{\mathbf{i}}_{\alpha I}^+$ from the image of the transmit antenna is reflected from the ground and bistatically scattered from the particle in the direction of $\hat{\mathbf{o}}_x = \hat{\mathbf{o}}_{\alpha I}^-$ toward the image of receive antenna after reflection from the ground. It involves double reflection from the ground.

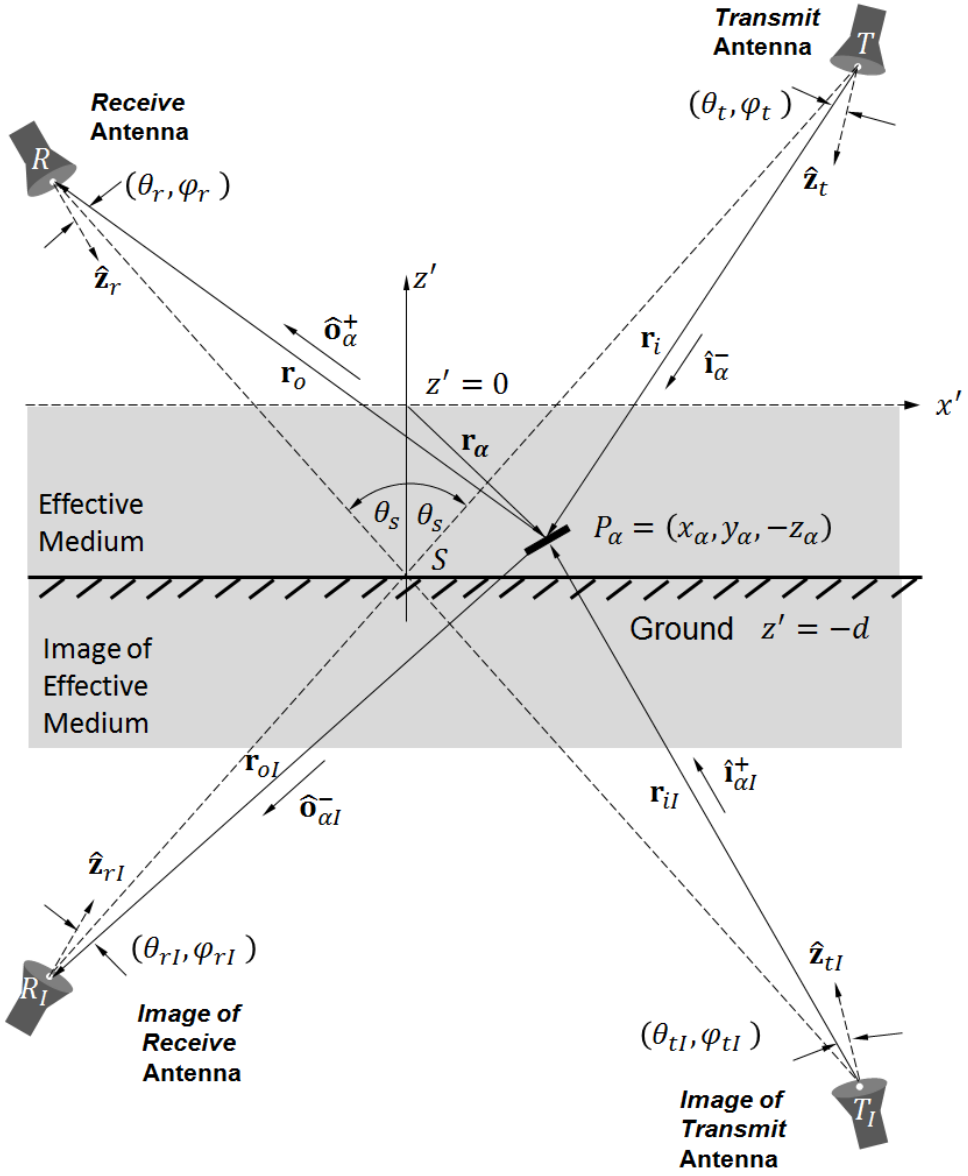

Figure 2.4: Scattering mechanisms and vector definitions for bistatic antenna configuration.

The bistatic scattering from the particle of type α is described by a bistatic scattering amplitude matrix $\underline{\underline{f}}_{\alpha}\left(\hat{\mathbf{o}}_x, \hat{\mathbf{i}}_x; \beta_{\alpha,n}\right)$ as

$$\underline{\underline{f}}_{\alpha}\left(\hat{\mathbf{o}}_x, \hat{\mathbf{i}}_x; \beta_{\alpha,n}\right) = \begin{bmatrix} f^{\alpha}_{\mathrm{VV}}\left(\hat{\mathbf{o}}_x, \hat{\mathbf{i}}_x; \beta_{\alpha,n}\right) & f^{\alpha}_{\mathrm{VH}}\left(\hat{\mathbf{o}}_x, \hat{\mathbf{i}}_x; \beta_{\alpha,n}\right) \\ f^{\alpha}_{\mathrm{HV}}\left(\hat{\mathbf{o}}_x, \hat{\mathbf{i}}_x; \beta_{\alpha,n}\right) & f^{\alpha}_{\mathrm{HH}}\left(\hat{\mathbf{o}}_x, \hat{\mathbf{i}}_x; \beta_{\alpha,n}\right) \end{bmatrix} \tag{2.15}$$

where the bistatic amplitude $f^{\alpha}_{\mathrm{pq}}\left(\hat{\mathbf{o}}_x, \hat{\mathbf{i}}_x; \beta_{\alpha,n}\right)$ is defined above. The quantities p and q denote the local polarization states along incident ($\hat{\mathbf{i}}_x$) and scattering ($\hat{\mathbf{o}}_x$) directions, respectively, and can be horizontal (H) or vertical (V). As stated above, the quantity $\beta_{\alpha,n}$ describes the orientation, size, and position statistics of the scatterer type. Note that the polarization of scattering amplitude is given in terms of linear polarization since polarization in the local ground frame is defined with linear polarization basis only.

The ground reflection matrix $\underline{\underline{r}}_g$ is defined in (2.9) and the transmission matrix t for the particle located at $\mathbf{r}_\alpha$ (the point P_α) is given by

$$\underline{\underline{t}}(\hat{\mathbf{i}}) = \begin{bmatrix} e^{i\Delta\kappa_{zp}(\hat{\mathbf{i}})r_{z\alpha}(\hat{\mathbf{i}})} & 0 \\ 0 & e^{i\Delta\kappa_{zq}(\hat{\mathbf{i}})r_{z\alpha}(\hat{\mathbf{i}})} \end{bmatrix} \tag{2.16}$$

where $\Delta\kappa_{zp}(\hat{\mathbf{i}})$ is defined in (2.8b) and $r_{z\alpha} = |\mathbf{r}_\alpha \cdot \hat{\mathbf{n}}|$ is the vertical distance that the wave propagates within the mean medium. For direct paths from and to the particle (antennas in the real positions), they are given by $r_{z\alpha}(\hat{\mathbf{i}}^{-}_{\alpha}) = r_{z\alpha}(\hat{\mathbf{o}}^{+}_{\alpha}) = z_\alpha$, while for reflected paths from and to the particle (antennas in the image positions), they are given by $r_{z\alpha}(\hat{\mathbf{i}}^{+}_{\alpha I}) = r_{z\alpha}(\hat{\mathbf{o}}^{-}_{\alpha I}) = -z_\alpha + 2d$. Since each scattering mechanism follows a different path, the total path traveled will be different. The factors $B^{x}_{\alpha,n}$ in (2.12) that account for the spreading loss (r_{sr}/r_o) and path dependent phase terms for each mechanism are thus unique.

The magnitude of these ratios could deviate from 1 when the receiver is closer to vegetation. In other words, vegetation depth relative to the receiver height determines the significance of this factor. They are given by

$$B_{\alpha,n}^{\mathrm{dd}} = \left(\frac{e^{ik_0(r_i+r_o)}}{r_i r_o}\right) / \left(\frac{e^{ik_0(r_{st}+r_{sr})}}{r_{st} r_{sr}}\right) \tag{2.17a}$$

$$B_{\alpha,n}^{\mathrm{dr}} = \left(\frac{e^{ik_0(r_{iI}+r_o)}}{r_{iI} r_o}\right) / \left(\frac{e^{ik_0(r_{st}+r_{sr})}}{r_{st} r_{sr}}\right) \tag{2.17b}$$

$$B_{\alpha,n}^{\mathrm{rd}} = \left(\frac{e^{ik_0(r_i+r_{oI})}}{r_i r_{oI}}\right) / \left(\frac{e^{ik_0(r_{st}+r_{sr})}}{r_{st} r_{sr}}\right) \tag{2.17c}$$

$$B_{\alpha,n}^{\mathrm{rr}} = \left(\frac{e^{ik_0(r_{iI}+r_{oI})}}{r_{iI} r_{oI}}\right) / \left(\frac{e^{ik_0(r_{st}+r_{sr})}}{r_{st} r_{sr}}\right) \tag{2.17d}$$

Figure 2.5 illustrates the directions of incoming and outgoing waves at an individual particle (i.e., a vertical trunk). The polarization basis vectors, defined at the particle, differ for each direction due to the beam divergence. As a result, four distinct polarization rotation matrices have to be constructed. Following the procedure in Appendix B, the rotation matrices from the particle to the receiver (denoted by the subscript $p \rightarrow r$) are obtained as:

$$\underline{\underline{u}}_{p\rightarrow r}\left(\hat{\mathbf{o}}_\alpha^+\right) = \begin{bmatrix} \hat{\mathbf{v}}_o^+ \cdot \hat{\mathbf{u}}_{r1}^* & \hat{\mathbf{h}}_o^+ \cdot \hat{\mathbf{u}}_{r1}^* \\ \hat{\mathbf{v}}_o^+ \cdot \hat{\mathbf{u}}_{r2}^* & \hat{\mathbf{h}}_o^+ \cdot \hat{\mathbf{u}}_{r2}^* \end{bmatrix} \quad \text{for } x = \mathrm{dd}, \mathrm{dr} \tag{2.18a}$$

$$\underline{\underline{u}}_{p\to r}\left(\hat{\mathbf{o}}_{\alpha I}^{-}\right)=\begin{bmatrix}\hat{\mathbf{v}}_{oI}^{-}\cdot\hat{\mathbf{u}}_{r1}^{*} & \hat{\mathbf{h}}_{oI}^{-}\cdot\hat{\mathbf{u}}_{r1}^{*}\\ \hat{\mathbf{v}}_{oI}^{-}\cdot\hat{\mathbf{u}}_{r2}^{*} & \hat{\mathbf{h}}_{oI}^{-}\cdot\hat{\mathbf{u}}_{r2}^{*}\end{bmatrix}\quad \text{for } x=\mathrm{rd},\mathrm{rr} \tag{2.18b}$$

and the rotation matrices from the transmitter to the particle (denoted by the subscript $t\to p$) are obtained as

$$\underline{\underline{u}}_{t\to p}\left(\hat{\mathbf{i}}_{\alpha}^{-}\right)=\begin{bmatrix}\hat{\mathbf{u}}_{t1}\cdot\hat{\mathbf{v}}_{i}^{-*} & \hat{\mathbf{u}}_{t2}\cdot\hat{\mathbf{v}}_{i}^{-*}\\ \hat{\mathbf{u}}_{t1}\cdot\hat{\mathbf{h}}_{i}^{-*} & \hat{\mathbf{u}}_{t2}\cdot\hat{\mathbf{h}}_{i}^{-*}\end{bmatrix}\quad \text{for } x=\mathrm{dd},\ \mathrm{rd} \tag{2.18c}$$

$$\underline{\underline{u}}_{t\to p}\left(\hat{\mathbf{i}}_{\alpha I}^{+}\right)=\begin{bmatrix}\hat{\mathbf{u}}_{t1}\cdot\hat{\mathbf{v}}_{iI}^{+*}\hat{\mathbf{u}}_{t2}\cdot\hat{\mathbf{v}}_{iI}^{+*}\\ \hat{\mathbf{u}}_{t1}\cdot\hat{\mathbf{h}}_{iI}^{+*}\hat{\mathbf{u}}_{t2}\cdot\hat{\mathbf{h}}_{iI}^{+*}\end{bmatrix}\quad \text{for } x=\mathrm{dr},\mathrm{rr} \tag{2.18d}$$

The local horizontal (parallel to local $x'y'$ plane) and vertical polarizations for each direction are given by

$$\hat{\mathbf{h}}_{o}^{+}=\frac{\hat{\mathbf{n}}\times\hat{\mathbf{o}}_{\alpha}^{+}}{\left|\hat{\mathbf{n}}\times\hat{\mathbf{o}}_{\alpha}^{+}\right|}\text{ and }\hat{\mathbf{v}}_{o}^{+}=\hat{\mathbf{h}}_{o}^{+}\times\hat{\mathbf{o}}_{\alpha}^{+} \tag{2.19a}$$

$$\hat{\mathbf{h}}_{oI}^{-}=\frac{\hat{\mathbf{n}}\times\hat{\mathbf{o}}_{\alpha I}^{-}}{\left|\hat{\mathbf{n}}\times\hat{\mathbf{o}}_{\alpha I}^{-}\right|}\text{ and }\hat{\mathbf{v}}_{oI}^{-}=\hat{\mathbf{h}}_{oI}^{-}\times\hat{\mathbf{o}}_{\alpha}^{+} \tag{2.19b}$$

$$\hat{\mathbf{h}}_{i}^{-}=\frac{\hat{\mathbf{i}}_{\alpha}^{-}\times\hat{\mathbf{n}}}{\left|\hat{\mathbf{i}}_{\alpha}^{-}\times\hat{\mathbf{n}}\right|}\text{ and }\hat{\mathbf{v}}_{i}^{-}=\hat{\mathbf{i}}_{\alpha}^{-}\times\hat{\mathbf{h}}_{i}^{-} \tag{2.19c}$$

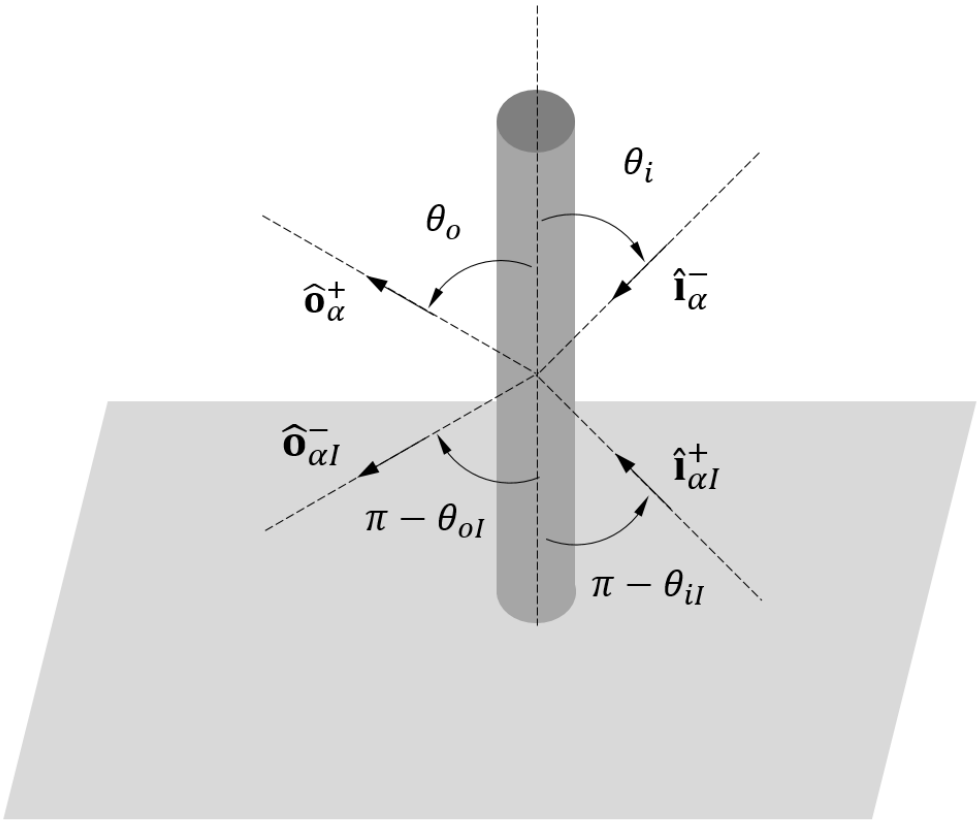

Figure 2.5: Unit scattering vector definitions.

$$\hat{\mathbf{h}}_{iI}^{+} = \frac{\hat{\mathbf{i}}_{\alpha I}^{+} \times \hat{\mathbf{n}}}{\left|\hat{\mathbf{i}}_{\alpha I}^{+} \times \hat{\mathbf{n}}\right|} \text{ and } \hat{\mathbf{v}}_{iI}^{+} = \hat{\mathbf{i}}_{\alpha I}^{+} \times \hat{\mathbf{h}}_{iI}^{+} \tag{2.19d}$$

4) Received Power: The total power received can be written in terms of the coherency vector as

$$\underline{P}_T = \underline{b} \otimes \underline{b}^* \tag{2.20}$$

where $\underline{b}^*$ is conjugate of b and the operator $\otimes$ indicates an outer product (Kronecker product) that is defined in Appendix D. Under the assumption that the scattering process is stationary and that the direct and multipath component fields arriving at the antennas from different directions are uncorrelated [72], the total power can be split into three independent terms as

$$\underline{P}_T = \underline{P}_d + \underline{P}_m^{\text{coh}} + \underline{P}_m^{\text{inc}} \tag{2.21a}$$

where

$$\underline{P}_d = \left\{ \frac{|K|^2}{r_d^2} \right\} \underline{D} \tag{2.21b}$$

$$\underline{P}_m^{\text{coh}} = \left\{ \frac{|K|^2}{\left(r_{st} + r_{sr}\right)^2} \right\} \underline{\Gamma}_s \tag{2.21c}$$

$$\underline{P}_m^{\text{inc}} = \left\{ \frac{|K|^2 A_S}{4\pi r_{st}^2 r_{sr}^2} \right\} \underline{\sigma}_e^0 \tag{2.21d}$$

where

$$\underline{D} = \underline{\underline{G}}_r \left(\hat{\mathbf{k}}_d \right) \cdot \underline{\underline{U}}_{t \to r} \left(\hat{\mathbf{k}}_d \right) \cdot \underline{\underline{G}}_t \left(\hat{\mathbf{k}}_d \right) \cdot \underline{E}_t \tag{2.22a}$$

$$\underline{\Gamma}_s = \underline{\underline{G}}_r \left(\hat{\mathbf{o}}_s^+ \right) \cdot \underline{\underline{U}}_{s \to r} \left(\hat{\mathbf{o}}_s^+ \right) \cdot \underline{\underline{R}}_s \left(\hat{\mathbf{o}}_s^+, \hat{\mathbf{i}}_s^- \right) \cdot \underline{\underline{U}}_{t \to s} \left(\hat{\mathbf{i}}_s^- \right) \cdot \underline{\underline{G}}_t \left(\hat{\mathbf{i}}_s^- \right) \cdot \underline{E}_t \tag{2.22b}$$

$$\underline{\sigma}_e^0 = \frac{1}{A_S} \sum_{\alpha} \sum_{n=1}^{N_a} \sum_{x} \left| B_{\alpha,n}^x \right|^2 \times \left[\underline{\underline{G}}_r \left(\hat{\mathbf{o}}_x \right) \cdot \underline{\underline{U}}_{p \to r}^x \left(\hat{\mathbf{o}}_x \right) \cdot \underline{\underline{\sigma}}_{\alpha,n}^x \left(\hat{\mathbf{o}}_x, \hat{\mathbf{i}}_x \right) \cdot \underline{\underline{U}}_{t \to p}^x \left(\hat{\mathbf{i}}_x \right) \cdot \underline{\underline{G}}_t \left(\hat{\mathbf{i}}_x \right) \cdot \underline{E}_t \right] \tag{2.22c}$$

where four-by-one vectors, such as the coherency vector $\underline{P}_T$, are written using uppercase letters and a single underline while four-by-four matrices are written using uppercase letters and a double underline. The coherency vectors $\underline{D}, \underline{\Gamma}_s$, and $\underline{\sigma}_e^0$ represent direct signal, specular reflectivity, and "effective" normalized bistatic radar cross section (NBRCS), respectively. Note that antenna effects (e.g., the antenna beamwidths, polarization crosstalk, loss, and altitude) are incorporated into NBRCS (or bistatic scattering coefficient), which should represent the object scattering properties only. The expression in (2.22c) is thus called "effective" NBRCS or $\underline{\sigma}_e^0$ and also indicated with subscript e. The effects of system parameters on NBRCS will be examined in the simulation section.

In calculating the outer products, the following identity is utilized $(\underline{\underline{m}} \cdot \underline{\underline{n}}) \otimes (\underline{\underline{m}} \cdot \underline{\underline{n}})^* = (\underline{\underline{m}} \otimes \underline{\underline{m}}^*) \cdot (\underline{\underline{n}} \otimes \underline{\underline{n}}^*)$ and the following vectors/matrices are defined.

The coherency vector of transmit antenna polarization state is given by

$$\underline{E}_t = \underline{e}_t \otimes \underline{e}_t^* \tag{2.23}$$

The antenna normalized (power) pattern matrices are given for the receive and transmit antennas, respectively, by

$$\underline{\underline{G}}_r = \underline{\underline{g}}_r \otimes \underline{\underline{g}}_r^* \tag{2.24a}$$

$$\underline{\underline{G}}_t = \underline{\underline{g}}_t \otimes \underline{\underline{g}}_t^* \tag{2.24b}$$

The polarization basis rotation matrices that operate on intensities are given by

$$\underline{\underline{U}}_{t\to r} = \underline{\underline{u}}_{t\to r} \otimes \underline{\underline{u}}_{t\to r}^* \tag{2.25a}$$

$$\underline{\underline{U}}_{s\to r} = \underline{\underline{u}}_{s\to r} \otimes \underline{\underline{u}}_{s\to r}^* \tag{2.25b}$$

$$\underline{\underline{U}}_{t\to s} = \underline{\underline{u}}_{t\to s} \otimes \underline{\underline{u}}_{t\to s}^* \tag{2.25c}$$

$$\underline{\underline{U}}^{x}_{p\to r} = \underline{\underline{u}}^{x}_{p\to r} \otimes \underline{\underline{u}}^{x*}_{p\to r} \tag{2.25d}$$

$$\underline{\underline{U}}^{x}_{t\to p} = \underline{\underline{u}}^{x}_{t\to p} \otimes \underline{\underline{u}}^{x*}_{t\to p} \tag{2.25e}$$

In (2.22b), the specular reflectivity operator that describes the scattering from the specular point is defined as

$$\underline{\underline{R}}_{S} = \underline{\underline{r}}_{S} \otimes \underline{\underline{r}}^{*}_{S} \tag{2.26}$$

In (2.22c), A_S is the surface area that is limited to a few Fresnel zones to capture most of the contributions and may differ from the whole area projected by the receiver antenna. In space borne geometries, the relative area (discriminated by both Doppler and Delay) are small and system parameters are uniform within that area. However, as demonstrated in Section 2.3-B, as the receiver gets closer to the ground (low altitudes), the system parameters and size of the surface area get significant in NBRCS calculations. The formulation in (2.22c) provides explicit expressions to correct the measured NBRCS by a ground-based system from the system-induced effects to be used in space-borne data interpretation.

Finally, the bistatic scattering cross-section operator is given by

$$\underline{\underline{\sigma}}^{x}_{\alpha,n} = 4\pi \left[\underline{\underline{s}}^{x}_{\alpha,n} \otimes \underline{\underline{s}}^{x*}_{\alpha,n} \right] \tag{2.27}$$

B. Special Case ($r_{st} \gg r_{sr}$) The expressions given above [namely, (2.21)] do not impose any restriction on the relative distance between receiver and transmitter slant ranges

except that both transmitter and receiver are in the far field of each other and both of them are in the far-zone of vegetation constituents above ground. On the other hand, in SoOp studies for remote sensing of earth properties, the utilized transmitters are mostly located in medium earth orbit (MEO) such as GNSS or geostationary orbit (GEO) such as communication satellites, and the receivers are operated from different platforms: ground, airborne, or spaceborne at low earth orbit. For all of these scenarios, the relative distance between transmitter and receivers to the specular point is large, i.e., $r_{st} \gg r_{sr}$. By taking advantage of this large difference, a flat earth approximation around the specular point can be applied. Moreover, for the transmitters at MEO and GEO, the range from the transmitter to the ground and the incidence angle direction can be assumed to be equal for all the surface surrounding the specular point. The footprint associated with Fresnel zones can be approximated with the following semi-minor and semi-major axes, respectively, [81]:

$$b = \frac{\sqrt{2\delta_n h_r \cos\theta_s}}{\cos\theta_s} \tag{2.28a}$$

$$a = \frac{b}{\cos\theta_s} \tag{2.28b}$$

where θ_s is the angle of reflection at the specular point as defined in (2.7), h_r is the receiver height with respect to ground, $\delta_n = n\lambda_0/2$ defines iso-delay ellipsis with the delay of multiples of half-wavelengths, and n is the Fresnel zone number. The n^{th} Fresnel zone is defined as the surface between the n^{th} and $(n-1)^{th}$ ellipsis. Even though one can use the exact expression for Fresnel ellipses given in [83], the above expressions remain fairly

accurate for most practical cases [67] and provide a better physical grasp of the footprint. The reflected GNSS signal mainly comes within the first ($n = 1$) Fresnel zone over land unless topographic effects are present as indicated in [163] while the reflected signals over ocean emanate over the multiple Fresnel zones (Glistening zone) [81]. It is important to identify contributing Fresnel zones and the scattering mechanism not only to determine the footprint but also to arrive at a physical model that is suitable for inversion.

The reflected signals are generally a superposition of both the coherent (specular), incoherent (diffuse) scattering, and direct components as given in (2.21). For down-looking antennas, the direct signal can be significantly suppressed, but the question remains on the relative importance of coherent and incoherent contributions over land with respect to platform altitude. It is important to recognize the path loss dependence with distance of each term. The specularly reflected signal follow a total ray path dependence, i.e., $1/(r_{st}+r_{sr})^2$ due to the image theory, while the diffuse term exhibits a multiplicative dependence given by $1/(r_{st}^2 r_{sr}^2)$ due to the additional spreading loss, the incident signal experiences after reradiation from the vegetation scatterers. The diffuse term thus decreases more rapidly with distance than the specular component. These differences are manifested in the relative weights of these two components for different receiving platforms: ground, airborne, and spaceborne.

Let us assume that specular reflection is dominant over land and find out how much diffuse scattering alters the specular reflectivity as a function of altitude. By normalizing the total power $\underline{P}_T$ with $|K|^2/(r_{st} + r_{sr})^2$, approximating the surface area ($= \pi ab$) with

(2.28), and neglecting the direct contribution, the reflectivity of the total reflected signal $\underline{\Gamma}_S^{(T)}$ can be obtained

$$\underline{\Gamma}_S^{(T)} \cong \underline{\Gamma}_S + \left(\frac{\delta_n}{2h_r}\right)\underline{\sigma}_e^0 \tag{2.29}$$

where the specular reflectivity $\underline{\Gamma}_S$ and the effective NBRCS $\underline{\sigma}_e^0$ are given in (2.22). The second term represents the contribution to reflectivity of incoherent scattering from the first "n" Fresnel zones. It is clear that the relative contribution of the NBRCS (diffuse contribution) is inversely proportional to the receiver altitude (h_r). This indicates that the specular component contains a greater portion of the total energy and thus contributes more to the received power unless: 1) NBRCS is significantly larger than specular reflectivity; 2) more Fresnel zones contribute; 3) the receiver is located at close neighborhood of the ground; or 4) any combination of the above cases happens. Diffuse scattering, although weak at high altitude, can have significant influence on ground and airborne systems which are usually used as a testbed for algorithm development for space-borne missions. In addition, the reflected signals could still include significant incoherent scattering components and the relative importance of these contributions is expected to vary with respect to the scene properties around the specular point. For these reasons, the diffuse component should be included in the model for low-altitude platforms. A bistatic model based solely on image theory cannot account for diffusely scattering.

2.3 Numerical Results and Discussion

The SCoBi-Veg model considers single scattering so it is intrinsically valid at P-, L-, and S-bands. There are already many potential free illuminators within these frequency ranges that could be used for earth science applications. Furthermore, the model could be extended to exploit signals at other bands if multilayer soil for lower frequencies or multiple scattering for higher frequencies are included. The particular communication satellites of interests to the recent studies [78, 56, 177] are the Navy's Mobile Users Objective System (MUOS) operating with RHCP transmit signals at P-band frequencies (360-380 and 240-270 MHz). There are four MUOS satellites at geostationary altitudes ($\sim$32,810 km), providing global coverage, except near the polar cap. As a preliminary application of the model, we will limit ourselves to P-band (370 MHz). The purpose of the simulations is two fold: 1) to show physical relevance of the model behavior to the SoOp studies in general and 2) to provide insights into polarimetric aspects of specular and diffuse contributions at P-band for a tree canopy observed from platforms at various altitudes.

The model requires various soil and vegetation parameters as inputs. In this paper, we will use the *in situ* parameters collected in an active/passive soil moisture experiment in Maryland in 2006 [94]. The experimental site consisted of plots of planted stands of deciduous Paulownia trees, a fast-growing deciduous tree with broad leaves shown in Figure 2.6. The tree plot used in this paper had 92 trees in a 1089-m^2 area. The dry biomass was about 9 kg/m^2, while the woody volume and density were 185.8 m^3/ha and 477.6 kg/m^3, respectively. The diameter at breast height (DBH) ranged from 17 to 23 cm (average DBH = 19.4 cm). The tree heights were variable, on the order of 11-14 m (average height = 13

m). Detailed measurements of the size/angle distributions of the tree constituents (trunk, branches, and leaves), along with their densities, were made. The results from the canopy sampling and dielectric measurements are shown in Table 2.1. Clear boundaries within canopy and trunk layers were identified by visual inspection. Figure 2.6 (right) shows the distribution of vegetation components within the vegetation layers. The soil texture at the site was a loamy sand, consisting of 80% sand and 7% clay. The ground was flat with a relatively smooth surface, where the surface rms height was on the order of 0.5-1 cm.

Table 2.1: Canopy parameters from destructive sampling [94].

Constituents		**Average Parameters**				
Types	**Kinds**	**Length (cm)**	**Radius (cm)**	**Density (m^{-3})**	**Dielectric Constant**	**Orientation Uniform**
Trunks	T1	617.0	8.73	0.005	15.6+i3.8	Vertical
Primary	B1	187.0	4.30	0.016	12.0+i2.9	20° - 50°
Branches	B2	153.8	1.58	0.188	12.0+i2.9	10° - 60°
Secondary	B3	63.6	0.98	0.734	12.0+i2.9	0° - 90°
Branches	B4	48.1	0.45	1.933	12.0+i2.9	0° - 90°
Leaves	L1	**Thickness 0.012**	10.2	11.12	35.2+i5.3	0° - 90°

A. Simulation Setting

The model considers both specular and diffuse scattering by incorporating the statistical and physical properties of the terrain into the received signal. While the calculation of the specular term requires properties of the mean vegetation attenuation in one direction and soil parameters at a single point, the diffuse term involves contributions of various scattering mechanisms due to each particle within the illuminated volume. In this present

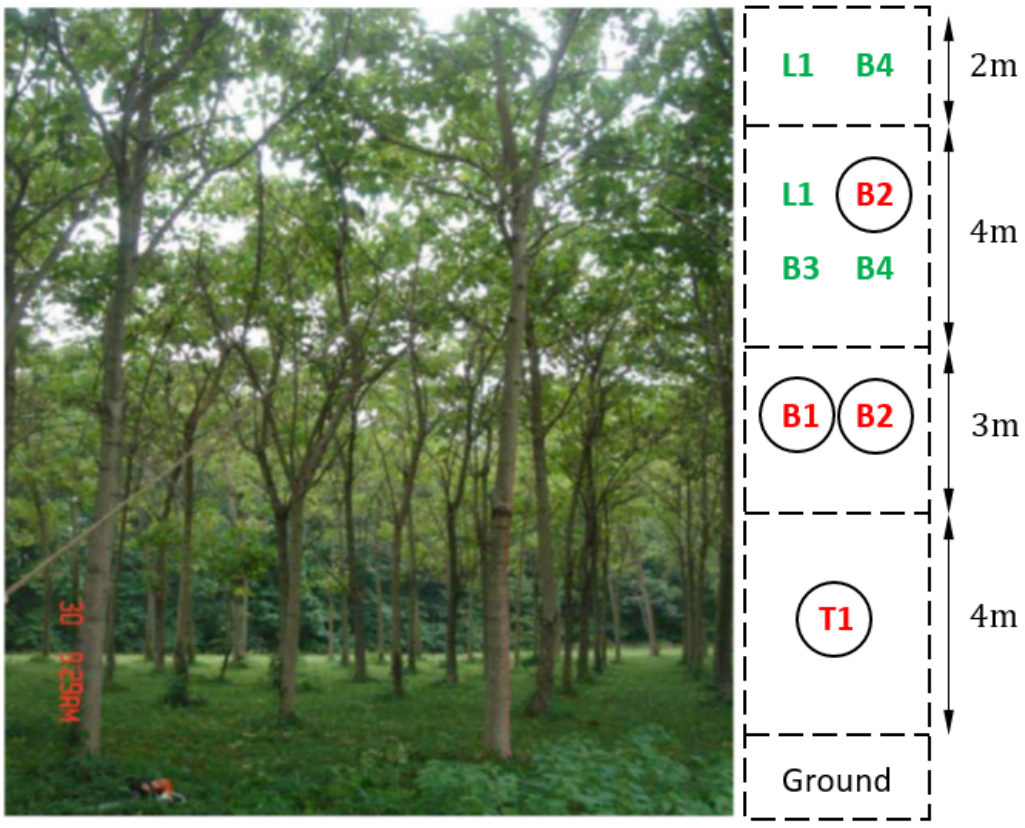

Figure 2.6: (Left) Paulownia stands [94]. (Right) Distributions of scatterers are illustrated. The scatterer kinds that are indicated by circles are primary branches or trunks that contribute the scattering while the others contribute the attenuation only. Canopy parameters are provided in Table 2.1.

investigation, a Monte Carlo procedure is chosen to predict the bistatic diffuse response from vegetation. Monte Carlo simulations are very useful to incorporate coherent addition and wave interaction effects in a vegetation canopy. The scheme holds possibility to incorporate realistic canopy structures such as architectural plant model [20], experimental vectorization [96], or recursive Lindenmayer systems [108] due to its generic model formulation. However, the base vegetation module in SCoBi-Veg is restricted for the moment to use a simple multilayer canopy model where the scatterers are spread uniformly between illuminated layers using a given distribution. Due to the single-scattering assumption, the scatterers are assumed to be independent and no mutual interactions are considered.

The scattering contributions of electrically large components only [e.g., primary branches (B1, B2) and trunks (T1)] are considered in order to speed up the computation while the calculation of the mean medium involves attenuation due to all scatterers. Scattering due to leaves (L1) and secondary branches (B3 and B4) is ignored since their dimensions are much smaller than the wavelength (81 cm) at P-band. The number of significant scatterers is calculated within the Fresnel zones based on their particle densities (see Table 2.1). They are then embedded in the equivalent (mean) medium where random positions are generated for these scatterers in the illuminated volume based on their associated layers as indicated in Figure 2.6; each scatterer is then oriented according to some prescribed orientation statistics given in Table 2.1. The solutions of the bistatic electric fields are computed for each particle by combining the configuration attributes, bistatic scattering amplitude from each significant scatterers, and ground reflection; these are then summed coherently. Finally, by repeating for many realizations of the random medium, the average

received field of the bistatic scattered field is obtained from the responses averaged over all realizations. We found that 20 realizations for 20-100-m altitudes and 10 realizations for 500-m altitude provided sufficient convergence in the diffuse term results. As the altitude increases, the required number of realization for good convergence gets reduced since the larger footprint size already adds randomness to the scattering at higher altitudes.

Over the years, various observing strategies have been developed for SoOp studies, mostly GNSS-R applications. These include the following:

1) geodetic ground-based nearly hemispherical RHCP antenna [97]; 2) horizontal-pointing vertically polarized antenna [134]; 3) down-looking LHCP antenna and an up-looking RHCP antenna [82]; 4) one RHCP up-looking antenna and two down-looking antennas with one RHCP polarized and the other LHCP polarized [39]; 5) a two-element dual linearly polarized patch array mounted side of an aircraft by steering a null to the direction of the undesired signal [56].

The model can handle any configuration stated above or more general configuration where vertically and horizontally polarized antennas for both the up- and down-looking. In the present investigation, we are, however, going to restrict the simulations to situation where the dual circularly polarized receiver antenna (both RHCP and LHCP) points directly to the specular point to understand the angular dependence of polarimetric reflected signatures. In other words, both transmitter and receiver are configured such that both are pointing to the specular point ($\theta_i = \theta_s$) for the simulations provided below. Furthermore, no Doppler frequency shifts are considered due to geostationary transmitters and low-altitude receiver assumptions considered in the present setup.

A generalized Gaussian antenna pattern, including 25-dB sidelobes, 25-dB crosstalk, and 30° beamwidth (arbitrary values), is assumed as an approximation to the radiation pattern needed in this formulation. This pattern will be assumed to be circularly symmetrical about the direction of peak intensity. In the following simulations, both a receiving antenna gain and EIRP equal to 0 dB were supposed. The actual values depend on the choice of illuminator, actual receiver antenna, and processing gain. They affect the received power in dB only by an offset and do not change its sensitivity to the reflected signature.

B. Results and Discussion

For most SoOp scenarios, it is assumed that only coherent scattering takes place over land, so the incoherent component is neglected. This assumption relies only on the specular scattering driven by the Fresnel reflection coefficients. On the other hand, the reflected signals are generally a combination of both the coherent (specular) and incoherent (diffuse) scattering components. In Figure 2.7, a comparison is presented between diffuse (the filled circle marker) and specular (the filled square marker) contributions over Paulownia trees as a function of observation angle (θ_s) at various receiver heights (20, 50, 100, and 500 m). Figure 2.7 (left) shows cross-polarized (RLRHCP transmit, LHCP receive) received power while Figure 2.7 (right) shows co-polarized (RRRHCP transmit, RHCP receive) received power. In Figure 2.7, the results of specular reflection from a bare soil are also added and are represented by filled diamond marker. As it can be seen, as the receiver altitude increases from 20 to 500 m, the diffuse term in both polarization is greatly reduced due to multiplicative dependence of path loss of the diffuse term [see (2.21d)] and the specular term remains the same since its path loss has a total path ray dependence [see (2.21c)] and

the receiver altitude is much smaller than the GEO satellite altitudes. Similar results were experimentally observed by a GNSS-R balloon experiment where the received power did not change with the balloon height [18], indicating that the coherent term dominates the received signal.

It is evident from the results in Figure 2.7 that the cross-polarized coherent term will always dominate the received power even for the receiver at low altitudes for flat terrains (i.e., in the absence of topographic relief) and cross-polarized diffuse term can thus be ignored. Similar result for cross-polarized reflected GNSS signals has been confirmed by spaceborne observations at L-band [163]. On the other hand, co-polarized diffuse term could be comparable or even larger than co-polarized specular term at lower incidence angles and low altitudes. As the incidence angle increases, co-polarized specular term starts to overpass the diffuse term again. This indicates that the reflected co-polarized signals could still include incoherent scattering components at low altitudes and the relative importance of these contributions is expected to vary with respect to the scene properties around the specular point and incidence angle. This has important implications for studies that use the co- and cross-polarization power ratio [39] to cancel out common factors and to make soil moisture inversion more robust against roughness. The diffuse contributions for low-altitude observations should be included in the model and interpretation of the results since both specular and diffuse terms have different dependence on the path loss, polarization, and incidence angle.

The diffuse scattering includes volume scattering within canopy, double bounce effects due to soil and trunk/canopy interactions, as well as attenuation by the vegetation

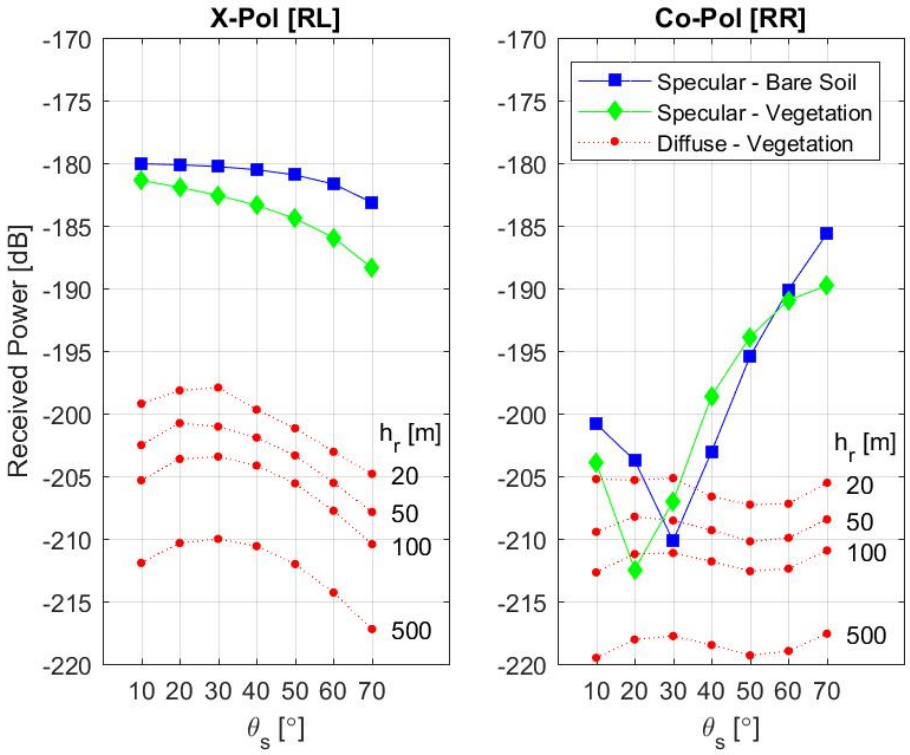

Figure 2.7: Relative contributions of coherent and incoherent components of received power as a function of incidence angle and the receiver altitude. (Left) cross-polarization (RL) received power. (Right) Co-polarization (RR) received power. Both receiver gain and EIRP were set to 0 dB. The received power will shift up with actual system EIRP, receiver gain, and processing gains (coherent/incoherent integrations).

canopy. Figure 2.8 provides comparison between specular reflectivity of bare soil (square marker) and vegetation (diamond marker), and effective NBRCSs (circular markers) of diffuse scattering mechanisms. It is well-known that ground reflection changes its circular polarization from right hand to left hand and vice versa. The cross-polarization specular reflectivity is thus much higher than co-polarization reflectivity at low incidence angles and both responses merge as the angle approaches Brewster angle as shown in Figure 2.8 (right). At lower incidence angles, the RR-polarized specular reflectivity first goes down and picks up again round 20°-30°. This behavior (same in Figure 2.7) happens due to the crosstalk (25 dB in this investigation) between the receiving antenna ports which makes a fraction of the signal go into the opposite polarization port. Furthermore, the crosspolarization specular responses of both bare soil and vegetation show a decreasing trend with increasing angle of observation, but the drop in vegetation reflectivity is more substantial due to the increasing attenuation by the plant canopy with increasing incidence angle.

In Figure 2.8 (left), the effective NBRCS is decomposed into its scattering mechanisms: single bounce, double bounce, and triple bounce. The single bounce shows no difference between cross- and co-polarized signatures since it involves no ground reflection and the volume scattering does not have any preference for circular polarization handiness. Similarly, the triple bounce shows no polarization difference either since the wave bounces twice from the ground. The first bounce changes RHCP to LHCP and the second bounce changes back to the RHCP. The triple bounce is also the smallest contribution due to double ground reflection. On the other hand, the double bounce response produces the largest among the other mechanisms since it bounces once from the ground and its scattering angle

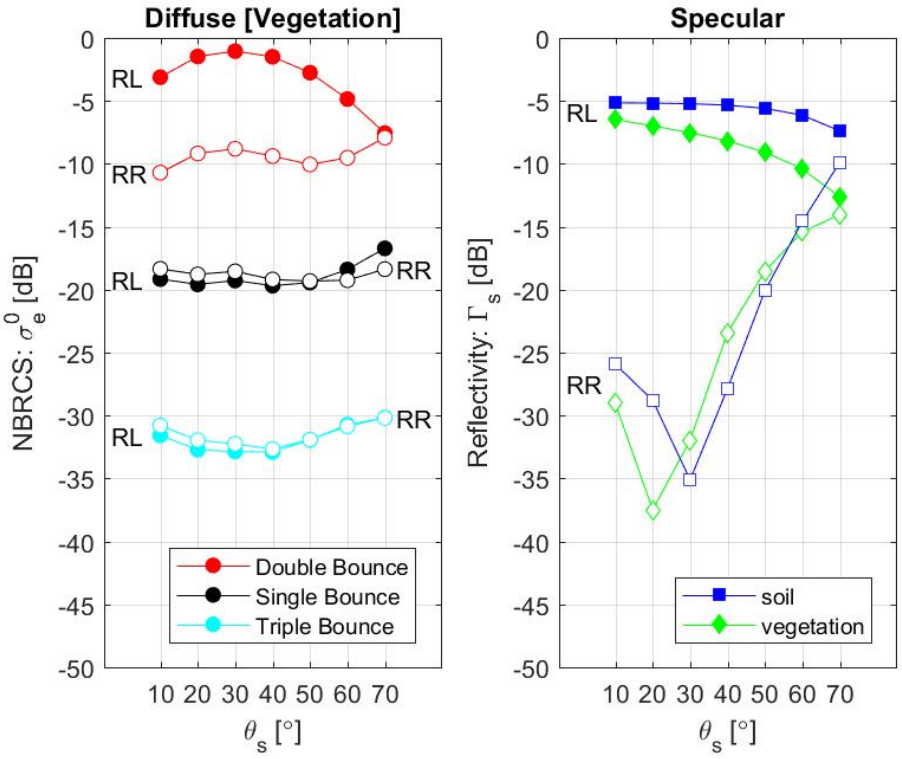

Figure 2.8: (Left) Angular responses of NBRCS scattering mechanisms and (Right) specular reflectivity. The filled markers denote cross-polarization (RL) while the unfilled markers represent co-polarization (RR). The receiver altitude is 500 m and the first Fresnel response is considered.

is well aligned with vertical trunks' scattering cones, where the amplitude of the scattering has its maximum in a conical region [31].

There are some remarkable features in both polarimetric and angular response of NBRCS. First, the cross-polarization NBRCS of the double-bounce contribution is higher than co-polarization NBRCS at low incidence angles and both response merges at higher angles. This is similar to the specular reflectivity response as described above, but the polarization difference is much smaller in NBRCS since volume scattering within vegetation combines two polarized components and reduces their relative strength. This difference is also much smaller than the present antenna crosstalk, which is 25 dB. This indicates that polarization mixing due to vegetation scattering is the main cause of this behavior. Second, cross-polarization NBRCS response shows an increasing trend up to $\sim 30^{\circ}$ incidence angle, and then decreases with increase in the incidence angle in contrast to the cross-polarization specular reflectivity that always decreases with the incidence angle.

Note that the both specular reflectivity and NBRCS of the vegetation depend not only on the surface and vegetation parameters but also on system parameters such as the receiving and transmitting antenna parameters, and altitude. In Figure 2.9, the effect of the receiver altitude on the diffuse term is further investigated by including the antenna characteristics and Fresnel zones by using expression in (2.22c) that provides the explicit dependence of NBRCS on the system parameters. In Figure 2.9(a), effective NBRCS is plotted against the number of Fresnel zones at various heights for ideal (nonfilled markers) and beam-limited (filled markers) antennas. The left panel shows cross-polarized (RL) NBRCS while the right panel shows co-polarized (RR) NBRCS. As can be seen from Fig-

ure 2.9, it is clear that: 1) in both polarization, effective NBRCS decreases as more Fresnel zones are included; 2) higher effective NBRCS is observed with higher altitudes; and 3) at lower altitudes, antenna radiation pattern plays a greater role in NBRCS values. It can be clearly noticed that NBRCS, which should represent the scene scattering properties only, is blended with the system parameters, particularly at lower altitudes. Most of the current experimental studies are carried out from lower altitude platforms that are often used before spaceborne implementation. In order to link these studies to spaceborne observation, the system-related effects need to be corrected to arrive at scene-only NBRCS.

Figure 2.9(b)(d) shows the projection of the receiver antenna footprints and 10 Fresnel zones on the ground from 20, 50, 100, and 500 m observing altitudes to provide further insight into the system-related effects on NBRCS. As evident from Figure 2.9(b)(d), the relative area covered by the same number of Fresnel zones with respect to antenna footprint gets reduced as the altitude increases. As a result, the scattering angles are ore confined around the boresight direction of the receiver antenna radiation pattern at higher altitudes. For instance, less decrease in NBRCS is observed at 500 m. When the results by using ideal (unit pattern matrix) and actual (beam-limited) antennas are compared, it is clear that even if we use pencil beam (ideal) antenna, the NBCRS still gets reduced as new Fresnel zones are included. This is due to the trunk's unique scattering patterns, which scatter most in a conical region. The scattering angles are much more spread at lower altitudes as shown in Figure 2.9(b)(d). Due to the spread of the wave, different particles in the layer experience different scattering directions. The weaker scattered fields are observed in the bistatic direction associated with trunk-ground interaction since the scattered wave from vertical

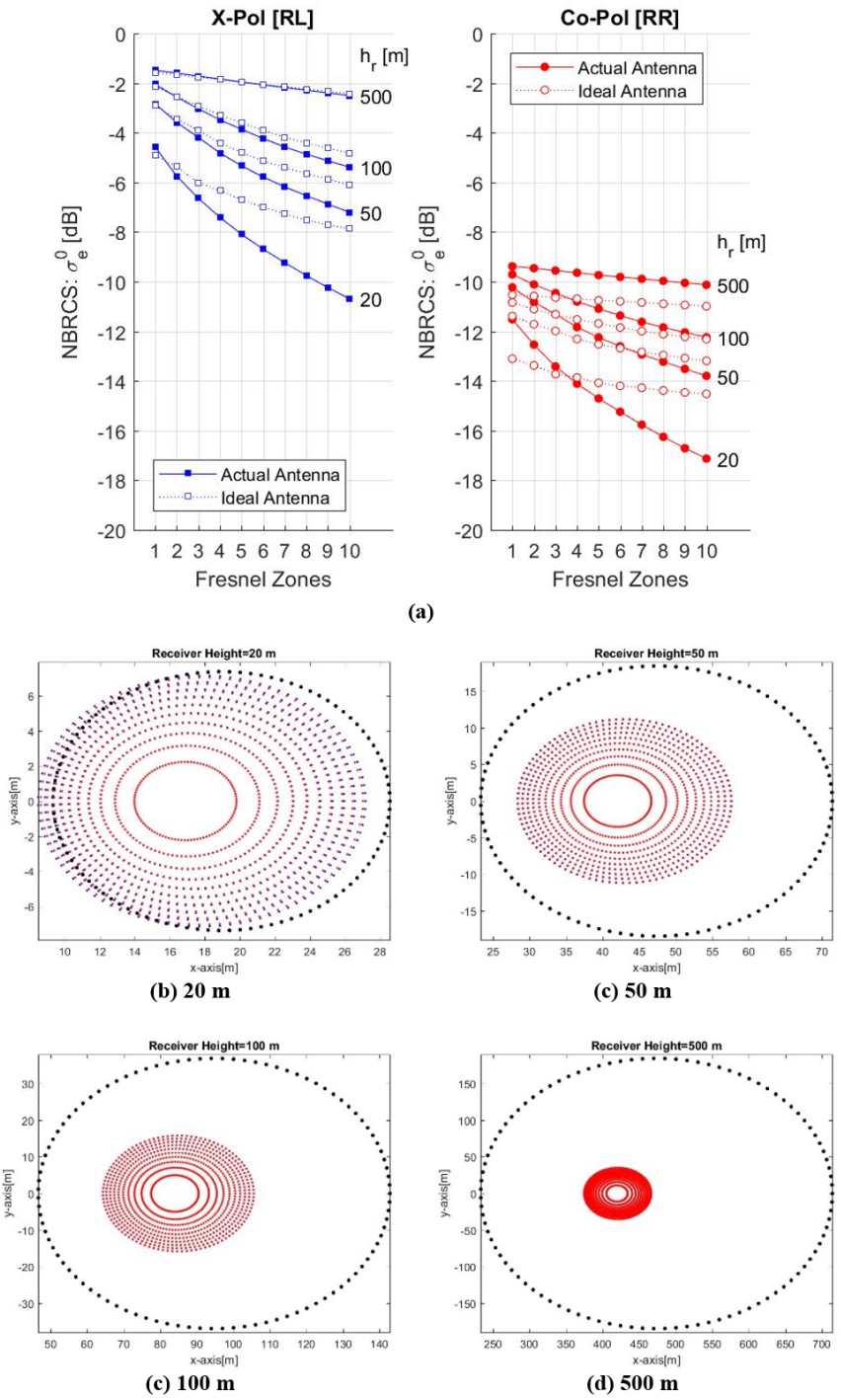

Figure 2.9: (a) "Effective" NBRCS as a function of increasing Fresnel zones (from 1 to 10), which is cumulative and includes inner Fresnel zones at various heights for ideal (the nonfilled markers) and beam-limited (the filled markers) antennas. (b)-(e) Fresnel zones and projected antenna footprint on the ground for receivers at various heights. The red ellipses (total of 10) are Fresnel zones while the black ellipse is antenna footprint.

trunks becomes slightly away from the forward scattering cone as shown in Figure 2.10(a). The amplitude of the scattering has its maximum in a conical region as illustrated in Figure 2.9(b). It is sometimes called skirt as well. For electrically large cylinders, this cone is sharp. As the scattering direction moves away from specular/forward direction, the specular scattering gets reduced drastically. This scattering behavior of vertical trunks acts like a special filter and concentrates most of the power along the specular direction since the local scattering angles get diverted from the specular directions for those that are located in higher Fresnel zones.

As stated earlier, the cross-polarization coherent scattering from vegetation over a flat ground almost always overpass the diffuse contribution to the received power. However, co-polarization diffuse power can be comparable or can be even larger than co-polarization coherent received power. It is worth investigating the effect of diffuse contribution on the co-polarization specular reflectivity. Figure 2.11 shows the effect of diffuse contribution in the first Fresnel zone as a function of altitude (left) and at 20-m altitude as a function of contributing Fresnel zones (right). The filled square marker represents specular reflectivity [(2.22b)] while smaller filled circles denote reflectivity of total received signal (specular + diffuse) [see (2.29)]. As seen from Figure 2.11 (left), as the altitude increases (increasing altitude is indicated by an arrow in Figure 2.11), the total reflectivity approaches the specular reflectivity. The effect of diffuse scattering is more apparent at lower incidence angles. On the other hand, as more Fresnel zones are included (increasing Fresnel zones are indicated with an arrow in Figure 2.11), the total reflectivity diverges from the specular reflectivity, particularly at lower incidence angles. This increase in the total reflectivity

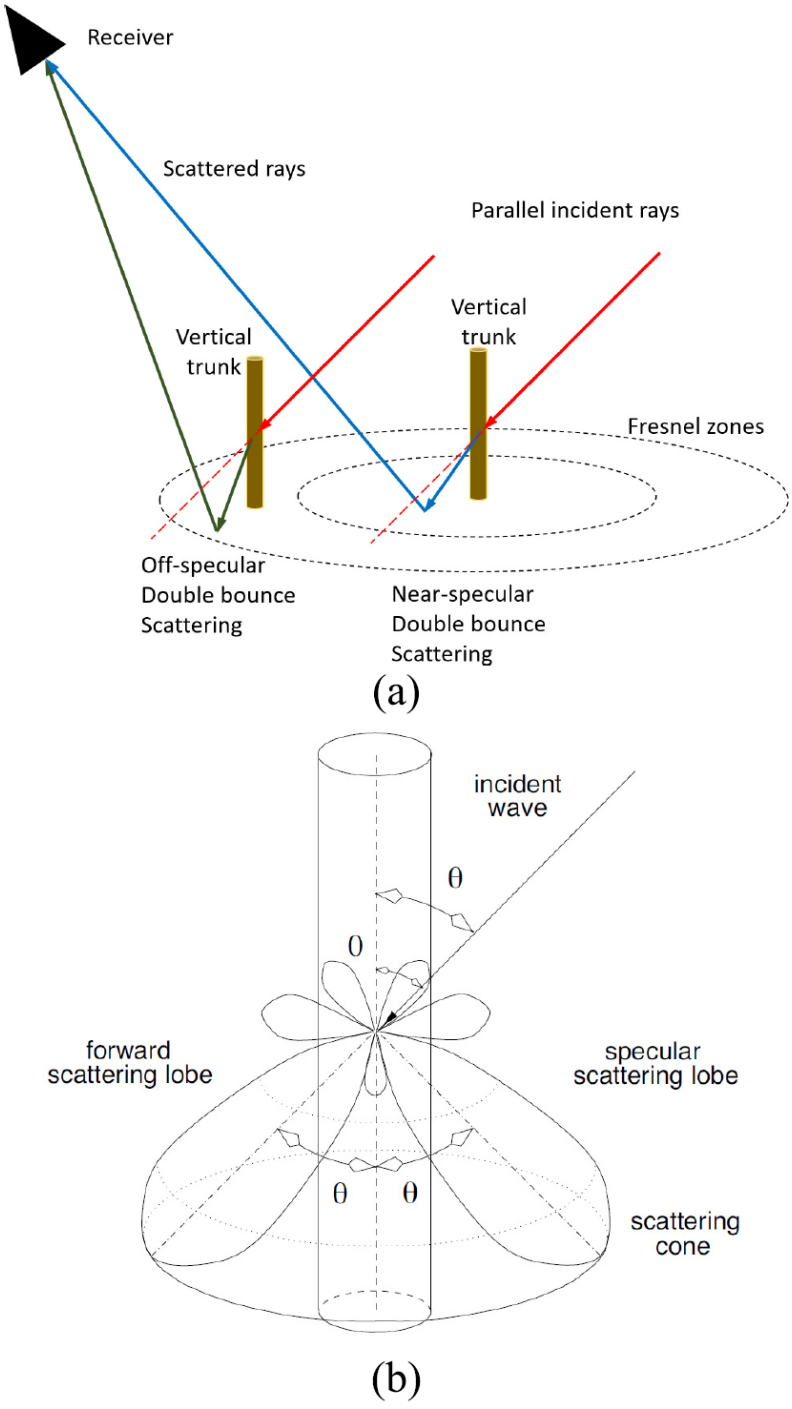

Figure 2.10: (a) Effect of beam divergence on double bounce. (b) Scattering pattern of a vertical trunk [31], [See also Figure 2.4].

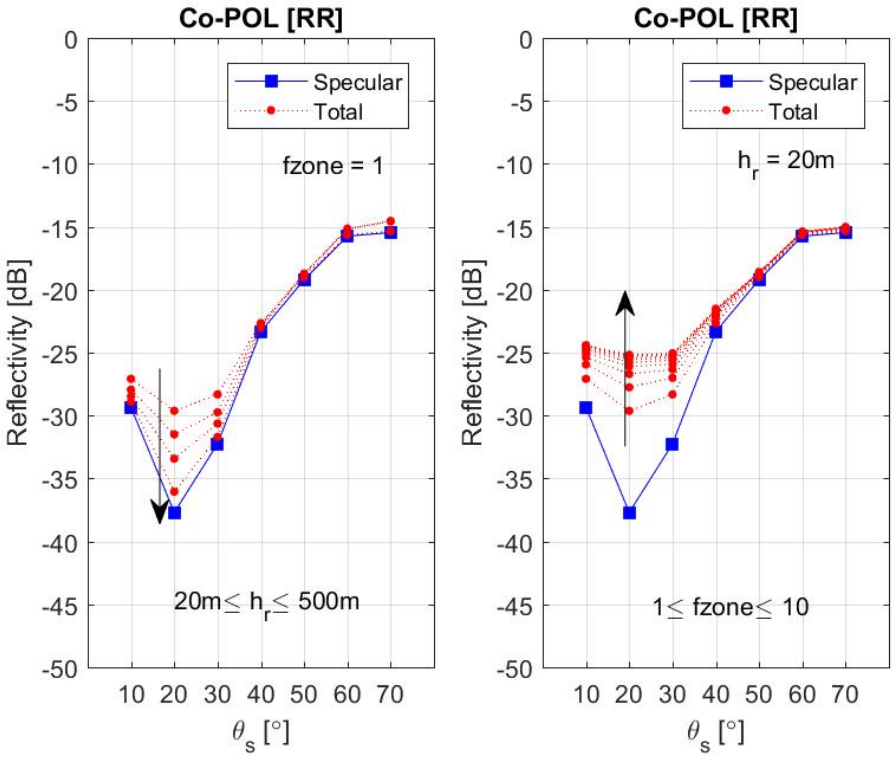

Figure 2.11: Effect of diffuse contribution on the reflectivity of total reflected signal. (Left) While the arrow indicates increasing altitude (20-500 m) for the first Fresnel zone, (Right) arrow indicates increasing Fresnel zone indices (1-0), which is cumulative and includes inner Fresnel zones, at a receiver height of 20 m.

is due to additional diffuse scattering with more Fresnel zones. As a result, additional diffuse contributions to the total reflectivity will alter its angular signature as a function of both Fresnel zones and receiver height and may introduce errors on the estimation of well-known Fresnel reflection coefficients that are often used in soil moisture retrievals. Diffuse scattering, although weak at high altitude, can have significant influence on co-polarization reflectivity observed by ground and airborne systems which are usually used as a test-bed for algorithm development for space-borne missions. If co-polarized signals are used in the retrievals, the co-polarized diffuse component should be included in the model for low-altitude platforms such as tower or small UAS.

2.4 Conclusion

There is increasing interest to use reflected (or multipath) navigation and communication satellite signals for remote sensing of a number of geophysical land parameters such as soil moisture and biomass. Although the handful experimental and theoretical studies demonstrate the potential of space-borne SoOp observations for such applications, there are many unknowns regarding the impacts of vegetation and system parameters on the observations. The scattering models are thus a critical means in advancement of these studies and in the design of future missions or field campaigns. In this paper, we presented an overview of the newly formulated coherent bistatic vegetation model, i.e., SCoBi-Veg, which simulates polarimetric reflectometry of vegetated landscapes using a Monte Carlo scheme. The model calculates the complex field in direct, specular, and diffuse terms by explicitly accounting for both system parameters (e.g., the antenna beamwidth, polarization

crosstalk, polarization mismatch, and altitude) and the statistical and physical properties of the terrain. In contrast to the previous studies that have usually assumed plane wave illumination/scattering, and/or ignored such antenna effects, the SCoBi-Veg model considers variations of both the strength and polarization states of the received wave along the beam direction so that the same model could be uniformly applied across different platforms at various altitudes.

P-band signals of geostationary communication satellites are considered as an application of the model. Although the results in this paper are not directly validated with experimental data and can be considered preliminary, simulated results provide insights into model's polarimetric and angular behavior with respect to system and scene parameters. To illustrate usefulness of the model in interpretation of field data, we present in-depth analysis of polarimetric specular and diffuse contributions to bistatic scattering from tree canopies at P-band. First, the coherent and diffuse components of the reflected signatures from tree canopies are compared for down-looking dual circular polarized receivers at various altitudes. It is shown that the cross-polarized coherent term dominates the received power over flat vegetated terrains, but co-polarized diffuse term could be comparable or even larger than co-polarized specular term at lower incidence angles and low altitudes. Second, the diffuse scattering was decomposed into the scattering mechanisms and the double bounce was shown to be the main contributor. Furthermore, the double bounce from the vertical trunks acts like a special filter and concentrates most of the reflected power along the specular direction. Third, the effect of antenna beamwidth and altitude on the diffuse scattering was also quantified. It is shown that the NBRCS, which depends on

the scene scattering properties only by definition, are blended with the system parameters, particularly at lower altitudes. Forth, even though the relative contribution of the NBRCS to the total reflectivity is inversely proportional to the receiver height, diffuse scattering can still alter angular signature of co-polarized reflectivity as a function of both Fresnel zones and receiver height. This may introduce errors on estimation of well-known Fresnel reflection coefficients that are often used in soil moisture retrievals. If co-polarized signals are used in the retrievals, the co-polarized diffuse component should be included in the model for low-altitude platforms and be corrected for antenna and altitude effects.

CHAPTER 3

SCOBI: A FREE, OPEN-SOURCE, SOOP COHERENT BISTATIC SCATTERING SIMULATOR FRAMEWORK

This chapter is inherited from the following publication:

O. Eroglu, D. Boyd, and M. Kurum, "SCoBi: A free, open-source, SoOp Coherent Bistatic Scattering Simulator framework," *IEEE Geoscience and Remote Sensing Magazine*, in press.

Significant author contributions to the above publication were as follows: Conceptualization and Methodology by **Eroglu** and Kurum; Investigation and Validation by **Eroglu**, Boyd, and Kurum; Writing by **Eroglu**; Review and Editing by **Eroglu**, Boyd, and Kurum; Visualization by **Eroglu**; Supervision by Kurum; Project Administration by **Eroglu** and Kurum.

Signals of Opportunity (SoOp) has the potential to offer cost-effective global remote sensing for land applications. Due to the complexity of SoOp scattering over land, comprehensive bistatic scattering models and simulators can help demonstrate its feasibility. To investigate this potential, we have developed a generalized, fully polarimetric forward model titled SoOp Coherent Bistatic Scattering Model (SCoBi). We have also developed a simulator framework, employing the SCoBi model to create an analysis environment for

a community of researchers, scientists, and users with little-or-no electromagnetic background. We aim to enable studying and analyzing new SoOp methods with varying configurations; determining the optimal cases for specific missions; generating, visualizing, and analyzing test data with the help of SCoBi. The SCoBi framework currently implements SoOp analyses over bare or vegetated terrains where soil can be modeled as single-layered or multi-layered dielectric medium. The software is free and open-source under the GNU General Public License (GPL) and is compatible with the MATLAB development environment. It obtains several inputs for the bistatic configuration (such as bistatic geometry, transmitter and receiver antenna characteristics, ground structure, and user preferences) through a set of user-friendly GUI windows. It generates the simulated direct received field and power as well as the specular reflection coefficient and reflectivity outputs. It also provides basic analysis functions such as plotting the reflectivity as a function of simulation parameters (such as transmitter's elevation angle or volumetric soil moisture). This paper describes the SCoBi simulator in detail and provides case studies for the Earth science community.

3.1 Introduction

Signals of Opportunity (SoOp) approach has been emerging in recent years because of its potential for global scale microwave remote sensing in Earth science applications. This approach is based on receiving the reflected signals of existing transmitters (such as navigation or communication satellites) that are scattered from the Earth surface in a bistatic geometry. Because a dedicated transmitter is not needed, SoOp enables high spa-

tiotemporal microwave remote sensing through the use of small satellite constellations. Depending on the geo-location of the specular point between the transmitter and the receiver, SoOp receivers can collect valuable information for both land- and ocean-based geophysical properties that is realizable through comprehensive inverse retrievals.

SoOp applications for the remote sensing of the ocean has seen advancements over the past two decades that have resulted in the launch of a few satellite missions. For instance, the first dedicated spaceborne GNSS-R receiver was a secondary payload on-board the UK Disaster Monitoring Constellation (DMC) [61]. It has demonstrated the potentiality of GNSS-R for the remote sensing of ocean, ice, and land geophysical parameters [60]. The UK Technology Demonstration Satellite (TDS-1) was launched in 2014 with an improved secondary GNSS-R payload to demonstrate SoOp's ocean remote sensing feasibility [54]. NASA's Cyclone Global Navigation Satellite System (CYGNSS) mission was launched in December, 2016, to improve weather predictions by estimating ocean winds between 38° south and 38° north latitudes [143, 141, 136]. CYGNSS has eight small satellites each with four channels in orbit, and it has a mean revisit time of seven hours. The constellation records a considerable amount of land observation data as well, which are publicly available. Nevertheless, SoOp land remote sensing has yet to demonstrate the same technology-readiness as ocean remote sensing. The major reason behind this is the lack of accessibility to bistatic land models that sufficiently account for the effects of many land geophysical parameters (such as surface and root-zone soil moisture, vegetation biomass, surface roughness, topography, and snow water equivalent) on SoOp observables.

Few studies have introduced their forward models to deal with bistatic geometries over bare-soil or vegetated terrains. These models can be grouped into two categories regarding their method: models that are based on either radiative transfer theory (originating from the law of energy conservation) or analytical wave theory (in conjunction with distorted Born approximation). The known radiative transfer models are Bi-MIMICS (Bistatic- Michigan Microwave Canopy Scattering) [106], an extension of Tor Vergata model [52], SAVERS (The Soil and Vegetation Reflection Simulator) that is another model employing Tor Vergata [130], and an extension of Bi-MIMICS [174, 175]. On the other hand, COBISMO (Coherent BistaticScattering Model) [155] is based on analtyical wave theory. Some of these studies utilize a previously developed monostatic model and extend it to the bistatic geometry [106, 174, 175], and another one exploits a monostatic coherent model [155], whereas the others are intrinsically bistatic. Each one of these studies considers diverse ground conditions and covers some of the microwave spectrum bands and polarizations. For instance, Liang et al. applied their model to tree canopies at L, C, and X-bands with linear polarization [106]. Ferrazzoli et al. performed forrest biomass analysis at L-band (GNSS-R) with circular polarization [52]. Pierdicca et al. studied bare-soil and vegetated terrains at L-band with circular polarization [130]. Two studies were applied to L-band (GNSS-R) with combinations of linear and circular polarizations [174, 175]. Thirion-Lefevre et al. performed P-band analysis with linear polarization [155]. In addition to these models, there is an open-source GPS multipath simulator [120] that is based on a coherent forward multipath model for near-surface reflectometry and positioning applica-

tions [119]. It is a comprehensive GPS analysis toolbox that can produce signal-to-noise (SNR) ratio, carrier phase, and code pseudorange observables.

Table 3.1: SCoBi model in a nut-shell.

Polarimetric effects	Fully polarimetric Combination of linear/circular polarizations
Antenna effects	Custom antenna patterns Cross-polarization coupling Beam divergence Polarization mixing Orientation
Configuration effects	Altitude Spreading loss over vegetation layer
Interferometric effects	Complex voltage Orientation Beam-forming
Multi-layer effects	Complex dielectric media Stratified layer division Vegetation and subsurface scattering
Vegetation effects	Mix vegetation Seasonal effects

We have recently developed a generalized, fully polarimetric, bistatic, forward model for the purpose of creating a comprehensive SoOp analysis environment. The model is originally entitled the SoOp Coherent Bistatic Scattering Model for Vegetated Terrains(SCoBi-Veg) [92, 93]. However, it is now applicable for vegetated and bare-soil terrains where the ground structure can be single-layered or multi-layered in either case. Hence, the model will be referred as SCoBi starting from this study. As its name suggests, it is a coherent model based on the analytical wave theory. The SCoBi model, as shown in Table 3.1, considers the antenna characteristics, polarimetric effects, and configuration influences espe-

cially because such models are commonly used for ground-based and low-altitude airborne simulations where these factors significantly impact the bistatic deliverables. For instance, it takes into account any combination of polarizations (linear and circular) in conjunction with the antenna effects such as polarization mismatch and crosstalk. It also incorporates the antenna altitude, orientation, and voltage pattern as well as interferometric effects such as complex voltage and beamforming. The SCoBi model applies analytical wave theory in conjunction with the distorted Born approximation to account for the vegetation layer. Hence, it allows the user to analyze the phase as well as the amplitude information through a single scattering assumption. The model is capable of simulating the complex field and received power (Modified Stokes vectors converted from the complex field quantities) constructed by three contributions: direct (the line of sight distance between the transmitter and the receiver), coherent (the shortest multipath distance in the specular direction), and diffuse (scattered waves from the entire scene due to the vegetation scatterers). The diffuse term is acquired through a Monte Carlo scheme that generates a sufficient number of realizations for canonical scatterers within the vegetation canopy. One can refer to [93] for an exhaustively theoretical explanation of the SCoBi model. SCoBi has been recently validated by P-band SoOp reflectometry measurements over land through simulations of measured field data collected at Purdue University's Agronomy Center for Research and Education (ACRE) during the 2017 growth season [58]. Bare-soil measurements during this season showed consistent results between measured and simulated refletivity values. In near future, the SCoBi model will be employed to cross-validate the measurements obtained by in-house built drone- and tower-based multi-frequency receievers as well as small

satellite missions such as CYGNSS [143] and Signals of Opportunity P-band Investigation (SNoOPI) [1].

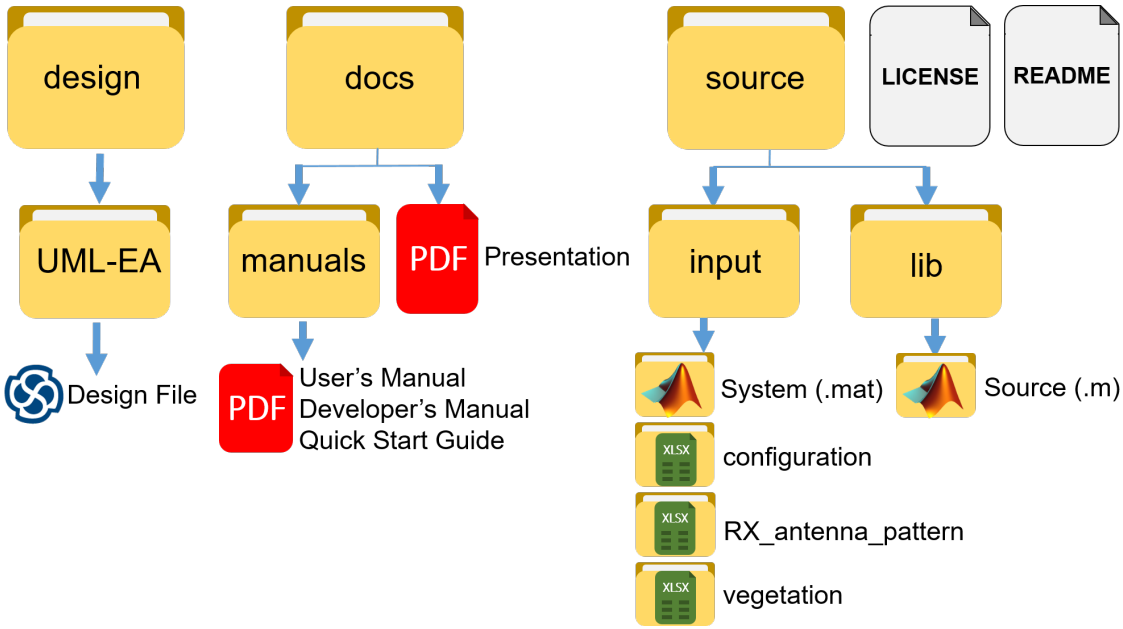

Figure 3.1: SCoBi Distribution Package. Icons represent individual files, and folder-enclosed icons represent folder names with several files.

This paper describes the free and open-source SCoBi simulator framework that we have further developed by employing the SCoBi model. The principal motivation for open-sourcing such a simulator is to create a medium for the Earth science community (including researchers, scientists, and even end-users with little-or-no electromagnetic background) to **(i)** study new SoOp methods under changing configurations, **(ii)** analyze the existing or newly-developed methods in detail, **(iii)** determine optimal cases for specific applications/missions, and **(iv)** generate, visualize, and analyze test data. From this point, SCoBi has the potential to contribute to the reproducibility of scientific efforts, technology val-

idations, and replicability of our simulation studies. SCoBi is also intended to optimize the time and effort spent by Earth science researchers on bistatic forward modeling needs. To acquire these objectives, the SCoBi simulator framework is released under the GNU General Public License (GPL).

The rest of the paper is organized as follows: Section 3.2 introduces the SCoBi simulator framework in detail, Section 3.3 demonstrates four case studies, Section 3.4 describes the design and implementation details of the framework, Section 3.5 discusses future inclusions within SCoBi, and Section 3.6 summarizes the conclusions.

3.2 SCoBi Simulator Framework

The distribution package of the SCoBi release can be accessed through GitHub [49] and is shown in Figure 3.1. The package consists of *design*, *docs*, and *source* folders as well as the license file (COPYING.txt) and a brief introduction (README.md) documentation. The *design* folder includes the architectural design of the SCoBi framework that is created by using the Sparx Systems' Enterprise Architect tool. The *docs* folder currently contains only the *manuals* folder that involves the SCoBi user's manual [48], developer's manual [46], and a quick start guide [47]. The *source* folder contains both the *input* and *lib* folders. The *input* folder includes the *system*, *configuration*, *Rx_antenna_pattern*, and *vegetation* folders. Each of these folders includes some number of default input files with the distribution. The details of these input files are described in the *Input Files* subsection.

This section will describe the features, available analysis types, inputs, and outputs of the SCoBi simulator framework. The simulator predicts fully polarimetric reflections

over vegetated or bare terrains where the ground is assumed as flat with a smooth to moderate surface roughness level. The ground structure can be modeled as single-layered or multi-layered slabs of dielectric constants for surface or root-zone analysis, respectively. Therefore, the framework currently provides bistatic analyses over vegetated and bare-soil terrains with surface-only (single-layered) or multi-layered ground structures. A major difference between the current simulator and the originally anounced SCoBi-Veg model [93] is that the diffuse scattering is not included in the current simulator. The decision to open-source the simulator without methods for calculating the diffuse scattering is an outcome of the significant findings from recent simulation studies [93, 51], which demonstrated the dominance of the coherent contributions over the incoherent scattering even for ground-based configurations in the case of either forested or agricultural terrain over a flat and smooth surface. For instance, terrains with a root mean square height (RMSH) roughness below 2.5 cm for L-band could generate well coherent reflections with respect to the Rayleigh roughness criterion. On the other hand, the SCoBi model is still capable of calculating the diffuse term despite not being implemented in the current simulator. Diffuse scattering feature will be added to the existing simulator in a future update for the purpose of applying SCoBi for terrains with topographic relief and high roughness levels.

The simulator calculates bistatic scattering for the most common SoOp configuration, which consists of a transmitter far-away from the Earth. The receiver model is a ground-based or low altitude passive instrument that acts as a reflectometer for the transmitted signals from long distances. A general bistatic geometry that SCoBi considers is depicted in Figure 3.2. The figure shows a satellite transmitter, a ground-based receiver, a multilayer

soil profile, and a vegetation layer. The local East-North-Up (ENU) coordinate system is depicted. In the figure, *S* represents the specular reflection point, and h_r and h_t represent the receiver and transmitter altitudes, respectively. In fact, h_r is not required for reflectivity analysis, but it can be given as an input for received power calculations. B_t and B_r represent the boresight (main beam) directions of the transmitter and receiver, respectively. The transmitter incidence angle (θ_{0t}) is defined as the angle between the transmitter's boresight direction and the nadir. Although this angle is used for bistatic geometry computations, SCoBi prompts the user to input the transmitter's elevation angle, which is complement of the incidence angle to 90°. The transmitter azimuth angle (ϕ_{0t}) is defined as the angle between the transmitter's local x-axis and local north (N). The receiver zenith observation (θ_{0r}) and azimuth observation (ϕ_{0r}) angles are defined similarly to incidence and azimuth angles of the transmitter, respectively. The defined azimuthal angles of the transmitter and receiver are converted into Cartesian coordinates to use in the computations after the input acquisition. The range term r_d denotes the line-of-sight distance between the transmitter and the receiver, r_{st} is the distance between the specular point and the transmitter, and r_{sr} stands for the distance between the specular point and the receiver.

The SCoBi simulator gets several inputs for the analysis type of interest and simulation specifications through a set of user-friendly graphical user interface (GUI) windows. It is applicable over a wide range of the microwave spectrum with any combination of linear and circular polarizations between the transmitter and receiver. It has a few constraints on the other hand. For instance, vegetation analysis is applicable from P to S-bands due to

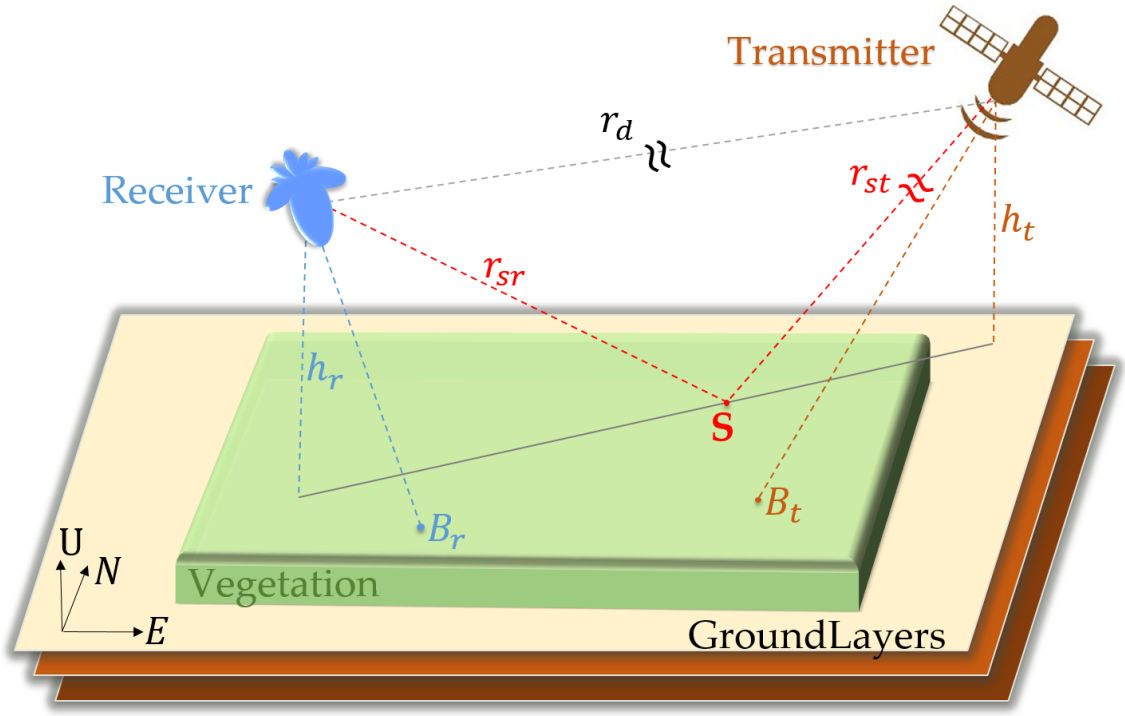

Figure 3.2: General bistatic geometry

the employment of the distorted Born approximation [93]. The application domain of the SCoBi simulator can be explained considering the analysis types and input sets as follows:

A. SCoBi Analysis Types

The SCoBi simulator has been developed as a framework in order to make it easily expandable for new analysis types. Currently, it provides vegetation and/or bare-soil analysis over single-layered or multi-layered ground structures. In addition, the ground model assume a flat smooth surface without topographic relief. SCoBi can be extended by adding other capabilities such as snow, topography, permafrost, and wetlands. Figure 3.3 shows the (Analysis Selection Window) of the SCoBi simulator, which is the initial window that welcomes the user when SCoBi is run by running the "runSCoBi.m" file under

the */source/lib* folder of the distribution package. This window has eight press-buttons for analysis types. The first four of these buttons are currently ready to use; however, the next four show how this framework can be extended in the future. For instance, SCoBi model can be further made capable to account for snow cover in conjunction with vegetation cover. In addition, the smooth ground assumption of the current model will be revised to consider the diffuse scattering over a highly rough ground so that topography analysis can be made possible. The currently active analysis selection buttons (*Forest, Agriculture, Soil*, and *Root-zone*) provide an easy way of selecting the SoOp analysis of interest and to prepare the simulator with default (provided for each analysis type in the software release) or recently-used inputs of that analysis type. In other words, the choice made in the *Analysis Selection Window* does not make any change in the use of SCoBi other than filling entries with the default or recently-used input values; the user should determine the appropriate input set for the analysis type selected.

SCoBi opens the *Simulation Inputs Window* after one of the analysis types is selected, as shown in Figure 3.4. This GUI window maintains the same layout for all the analysis types; however, the visibility of the GUI elements is managed based on the corresponding default or recently-used inputs.

B. SCoBi Inputs

SCoBi allows the user to determine a large number of input parameters to specify the bistatic geometry, scene characteristics, transmitter and receiver antenna specifications and orientations, and dynamic scene conditions. The simulator receives the simulation inputs with the help of a GUI window, *Simulation Inputs Window*, as seen in Figure 3.4. In

Figure 3.3: Analysis Selection Window of SCoBi

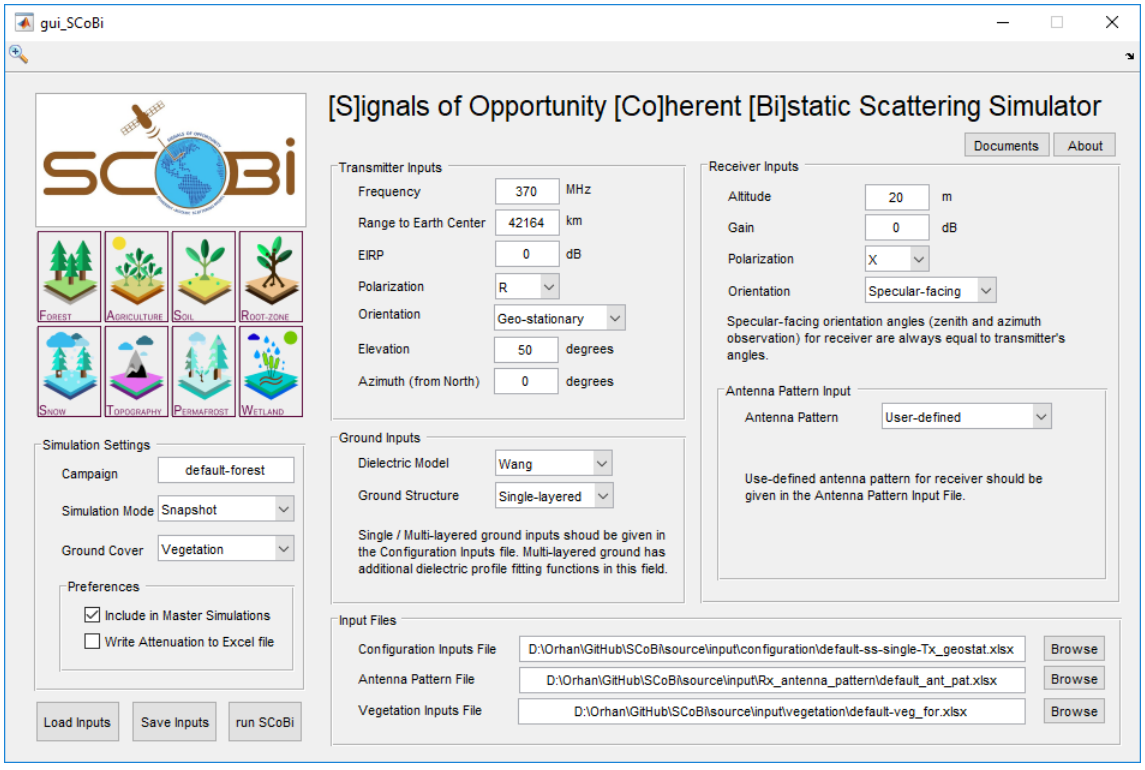

Figure 3.4: Simulation Inputs Window of SCoBi

fact, we have created a SCoBi user's manual [48] and a set of tutorial videos [9] to help visualize what the parameters within the *Simulation Inputs Window* represent and view how changing these parameters affect the input structure of the *Configuration Inputs File, Antenna Pattern File*, and *Vegetation Input File* that are discussed later in this paper. In addition, the separate input groups on the *Simulation Inputs Window* are briefly described as follows:

i. Analysis Selection Buttons

These buttons are included here for the sake of allowing the user to switch to load the default or recently-used inputs for the different analysis types at any time without returning to the *Analysis Selection Window*.

ii. Simulation Settings

The *Simulation Settings* panel includes the main setting parameters of a simulation that determine the output directory name, simulation mode, ground cover, and preferences. The *Campaign* field expects any character input and is combined with the timestamp of the simulation to generate the simulation output name. *Simulation Mode* is one of the most significant features of the SCoBi simulator, which enables *Snapshot* and *Time-series* analyses. The former is dedicated to simulate every combination of varying configuration and scene dynamics such as transmitter elevation and azimuth angles, volumetric soil moisture (VSM), and RMSH in order to easily generate large amounts of simulated datasets. The latter allows the user to run temporal analyses where every simulation iteration considers a corresponding set of the aforementioned dynamic parameters in conjunction with timestamps. The default input files use a Day-of-Year (DoY) numbering system. The simulation

mode selection affects the content of the *Configuration Inputs File*, which is described in the *Input Files* subsection. Ground cover allows the user to perform an analysis over a vegetated or bare-soil terrain. If it is selected to have vegetation cover, then SCoBi requires a *Vegetation Inputs File* from the user, which is described in the *Input Files* subsection. The SCoBi user can also decide on a few preferences such as whether including a simulation in a master logging option or performing it as temporary, and writing vegetation attenuation to an Excel file.

iii. Action Buttons

The *Simulation Inputs Window* houses three action buttons that are *Load Inputs, Save Inputs*, and *run SCoBi*. The first button is used for loading an existing system input (".mat" extension) into the *Simulation Inputs Window*. The second one can be used to save the current input parameter values on the *Simulation Inputs Window* as a SCoBi system input file with the ".mat" extension. The *run SCoBi* button allows the user to run a simulation with the inputs given in the *Simulation Inputs Window*.

iv. Transmitter Inputs

This panel consists of editable text fields and pop-up menus to determine the transmitter characteristics and orientation of the SCoBi simulations. The SCoBi user is able to input the transmitter operating frequency, range to Earth center, and Effective Isotropic Radiated Power (EIRP) from this panel. The operating frequency should be chosen appropriately from the microwave spectrum for the analysis requirements. The range should be determined regarding the SCoBi simulator's assumption that the transmitter is a far-away satellite from the Earth. The EIRP is included in the inputs in order to account for its offset

effect on the received power. The transmitter can be chosen to be right-hand circular polarization (RHCP), left-hand circular polarization (LHCP), X-polarized, or Y-polarized. In the SCoBi model, the y-axes of the antennas are considered parallel to the ground plane; thus, Y-polarization stands for the H-polarization (horizontal). X-polarization accounts for the V-polarization (vertical) since the x-axes of the antennas are complementary perpendicular to their y-axes. The *Orientation* of the transmitter allows the user to include either a geo-stationary or variable-orientation satellite. The orientation selection also makes changes in the preparation of the *Configuration Inputs File* such that the elevation and azimuth angles of a variable-orientation transmitter should be given in that file. On the other hand, the angles of a geo-stationary transmitter can be directly given in the *Simulation Inputs Window*. The boresight and specular directions of the transmitter are assumed to be the same in the SCoBi model because the transmitter is far-away from the Earth and these directions differ by a small angular difference. However, the system can be extended to input the real transmitter orbit dynamics in the future. Another potential future extension to SCoBi could account for geo-synchronous transmitter orientation as well.

v. Receiver Inputs

The content and purpose of this panel is similar to that of the transmitter inputs. The receiver *Altitude* and *Gain* are given as inputs for the power calculations. The polarization of the receiver can be chosen in the same way with the transmitter. The receiver Orientation can be specular-facing or fixed. The specular-facing orientation means that the main beam axis of the receiving antenna always faces the specular reflection point; thus, the receiver's zenith observation and azimuth observation angles are equal to the transmitter's incidence

and azimuth angles, respectively. This orientation is aimed at directly analyzing any SoOp configuration's angular sensitivity for varying incidence angles by minimizing the antenna pattern effects and polarization mismatch, which might significantly dominate the ground-based observations. The fixed orientation makes the receiver boresight direction always look at a fixed point, and fixed orientation angles (zenith and azimuth observation) can be given directly in the *Simulation Inputs Window*. These receiver orientations enable ground-based analysis; however, the SCoBi simulator can be extended for low-altitude airborne instruments by adding a variable receiver orientation in the same fashion that the transmitter inputs panel allows. In this case, the varying receiver orientation angles can be given in the *Configuration Inputs Files* like the transmitter so that, for example, time-series analysis from the observation of an airborne receiver with variable roll, pitch, and yaw can be performed. The last receiver input parameter is the *Antenna Pattern* selection. SCoBi currently provides two different antenna pattern specification methods: User-defined and Generalized-Gaussian. The user-defined antenna pattern definition is a powerful feature which allows the user to employ both simulated and measured antenna radiation patterns by simply providing the normalized voltage pattern values to the system (such that theta (θ) angles span 180°, while phi (ϕ) angles span 360°). The normalized voltage pattern of the antenna (co-polarization and cross-polarization for dual ports) may be measured in an anechoic chamber or generated in any antenna modeling tool and fed into SCoBi as the *Antenna Pattern File* as described in *Input Files* subsection. The second option to define the antenna pattern is the Generalized-Gaussian, which simply generates a 3D-symmetrical

antenna pattern with side-lobes. This option can be given its inputs, which are half-power beam-width, side-lobe levels, and cross-polarization levels, directly on the GUI window.

vi. Ground Inputs

This panel is used for deciding on the ground dielectric model and the layer structure. The SCoBi simulator offers the use of Dobson [35], Mironov [115], or Wang [170] dielectric models. The ground can be modeled as single-layered or multi-layered. The former provides a ground surface analysis only, whereas the latter is especially useful for root-zone analyses. The selection of the *Ground Structure* changes the content of the *Configuration Inputs File*, which is described in the *Input Files* subsection. SCoBi has an unprecedented strength because it allows to run both vegetation and bare-soil analysis over either single-layered or multi-layered ground structures. The rest of the ground inputs (such as soil texture or multi-layer discretization) are given through the *Configuration Inputs File*.

The *Simulation Inputs Window* shows four different dielectric profile generation methods when the ground structure is *Multi-layered*. The user can choose from one to all of these methods to generate dielectric profiles and to simulate GNSS-R observables (such as reflectivity) for these profiles. The Discrete-slab option can be selected to make SCoBi calculate the midpoint between each pair of VSM measurements within the soil layer. The user-defined soil moisture is then assumed to represent the soil moisture for the entire layer. For instance, for a VSM profile that is defined at points 5, 10, 20, and 40 cm within the soil moisture profile, the VSM at point 5cm will be representative of the calculated layer extending from 0 to 7.5 cm. The VSM value at 10 cm will represent values from 7.5 to 15 cm. The VSM value at 20 cm will represent values from 15 to 30 cm. The VSM value at

40 cm will represent the VSM from 30 cm to (40 cm + the bottom layer size). The bottom layer size (*zB*) is defined on the *Ground* sheet of the *Configuration Inputs File*, which is described in the *Input Files* part. The other available fit functions include 2^{nd} order polynomial, 3^{rd} order polynomial, and logistic regression fit. When any of these are selected, the value for *delZ*, which is defined on the *Ground* sheet of the *Configuration Inputs File*, is used to generate a series of points from the top of the soil moisture profile (z = 0) to the bottom.

vii. Input Files

The SCoBi simulator runs the simulations through the *Simulation Inputs Window*. The input parameters within this window are saved as a system input file with ".mat" extension whenever a simulation is run or the user saves the input. The default system input files (*Forest, Agriculture, Soil*, and *Root-zone*) are stored in the following directory (also shown in Fig.1):

$. \setminus source \setminus input \setminus system \setminus$

There are additional input files that are fed into the simulations, which are Microsoft Excel files. The need for these files and their contents may differ depending on the given inputs through the *Simulation Inputs Window*. A comprehensive visualization and explanation of these Excel input files and their interactions with SCoBi can be found in the user manual [48] and tutorial videos [9]. These Excel input files can be described as follows:

Configuration Inputs File: This file is required in every simulation and consists of two Excel spreadsheets: *Dynamic* and *Ground*. The *Dynamic* spreadsheet defines the variable parameters that may be given changing values within every simulation. The variable

parameters include timestamps (DoYs), azimuth and elevation angles of the transmitter, RMSH roughness, and VSM. The *Ground* spreadsheet includes the parameters that define the ground texture and multi-layering. The simulation mode, transmitter orientation, and ground structure selections determine the preparation of the content of this file. For example, the multilayered soil moisture profile is defined using both the *Dynamic* and *Ground* sheets of the *Configuration Inputs File*. On the *Ground* sheet, values for soil bulk density, sand ratio, clay ratio at multiple locations within the soil profile should be given, whereas the Single-layered ground case requires only one value for each of these parameters. Values for the air layer size (*zA*), bottom layer size (*zB*), and layer discretization (*delZ*) are also defined, while these are not included in a *Single-layered* simulation. Similar to values recorded by soil moisture data loggers, the *Dynamic* sheet lists the VSM at the different layer depths for a Multi-layered ground structure. When a time-series analysis is performed, each row is a different point in time as denoted by the *DoY* column. Thus, the *Dynamic* sheet lists the different VSM measurements as a function of time. For further advanced details, please refer to the user's manual [48]. In addition, a number of default *Configuration Inputs Files* comes with the SCoBi software distribution to help the user understand the configurations needed to prepare this file for different analysis types in the following directory of the SCoBi download package:

$. \setminus source \setminus input \setminus configuration \setminus$

Antenna Pattern File: This file is required only if the receiver antenna pattern is selected as *User-defined*. Otherwise, the *Simulation Inputs Window* disables the corresponding parts that allow to input this file. It consists of four spreadsheets: *gnXX, gnXY, gnYX*, and

gnYY, which holds the normalized voltage pattern values for co-polarization and cross-polarizations for dual ports. The default *Antenna Pattern File* that comes with the SCoBi software distribution can be found in the following directory of the downloaded package:

$. \setminus source \setminus input \setminus Rx\_antenna\_pattern \setminus$

Vegetation Inputs File: This file is required only if the *Ground cover* is selected as *Vegetation*. Otherwise, the *Simulation Inputs Window* disables the corresponding parts that allow the user to input this file. It consists of two spreadsheets: *Layers* and *Kinds*. The *Layers* spreadsheet should provide detailed information about the vegetation layer thicknesses and the content of each individual layer. Vegetation cover can be divided into a number of layers by defining separate layer thicknesses for the purpose of placing scatterers titled *Kinds (Leaf, Branch, Trunk, or Needle)* into different layers. For each vegetation layer row, the subsequent columns following the thickness column should define the content (involved constituent *Kinds*) of the layer specified by the row. The detailed information of these *Kinds* such as the dimensions, densities, orientation, and dielectric constants should be given in the *Kinds* spreadsheet. New *Vegetation Inputs File* other than the default ones can be prepared following the user's manual [48]. Default files for Forest and Agricultural analysis come with the SCoBi release under the following directory:

$. \setminus source \setminus input \setminus vegetation \setminus$ ***C. SCoBi Outputs***

The basic relationship between the inputs, SCoBi source-code (*lib* folder), and the outputs can be considered as the use of a black-box tool from the user's perspective, as shown in Figure 3.5. In other words, the SCoBi user prepares the inputs, run the simulator only interacting with the *runSCoBi.m* function, SCoBi uses the inputs (both system

input and Excel inputs), and generates the corresponding simulation outputs under the $./setminussource/setminussims$ directory. It generates the complex field and real valued power results for the directly received signals, and reflection coefficient and reflectivity values for the specular reflection contribution. It is also possible to calculate the received power due to the specular reflection contribution by using the SCoBi outputs. The sizes of the resulting matrices depend on the simulation inputs (snapshot or time-series simulation, and the number of varying values for dynamic system parameters).

The current SCoBi simulator version has a limited number of analysis functions such as plotting the reflectivity as a function of transmitter elevation angle or volumetric soil moisture. These initial plotting functions are provided to help user simply visualize the reflectivity; however, the simulator outputs are ready to be analyzed for phase or amplitude as well. Further details about the SCoBi outputs and analysis features can be found in the user's manual [48] and the tutorial videos [9].

Figure 3.5: The black-box relationship between the input, lib, and output folders from the user's perspective.

3.3 Case Studies

The default inputs of the system are aimed at allowing the user to easily learn the use of SCoBi system and differences between analyses. They enable performing four differ-

ent case studies with the initial distribution package as well. These are Forest, Agriculture, Soil, and Root-zone. The case studies are given only with the reflectivity results in this section; however, the bistatic phase, interferometry, reflection coefficient, or received power can be analyzed further by advanced users. Moreover, the effects of varying custom antenna patterns and specifications (polarization, polarization mismatch and crosstalk, orientation, and altitude) as well as different vegetation statistics on the reflectivity can be studied.

A. Forest

The default input for the Forest analysis, as seen in Figure 3.4, allows to make a P-band simulation over a forested terrain of deciduous Paulownia trees. It considers a geostationary communication satellite that is operating at P-band (370 MHz). It simulates a ground-based, X-polarized receiver that has an altitude of 20 meters. A single-layered ground structure is modeled in this forest case. The vegetation layer is designed to resemble a Paulownia forest by including trunks, branches, and leaves with the statistics of dimensions, densities, orientation, and dielectric characteristics coming from an in situ measurement data [94]. This example is described comprehensively within the SCoBi user manual [48]. The simulated reflectivity as a function of VSM values used in these simulations can be seen in Figure 3.6.

B. Agriculture

The default input for the Agriculture analysis, on the other hand, considers an agricultural field, where vegetation inputs enable modeling the corn constituents. The transmitter is a GNSS satellite that operates at L1A-band (1575.42 MHz). The ground structure and

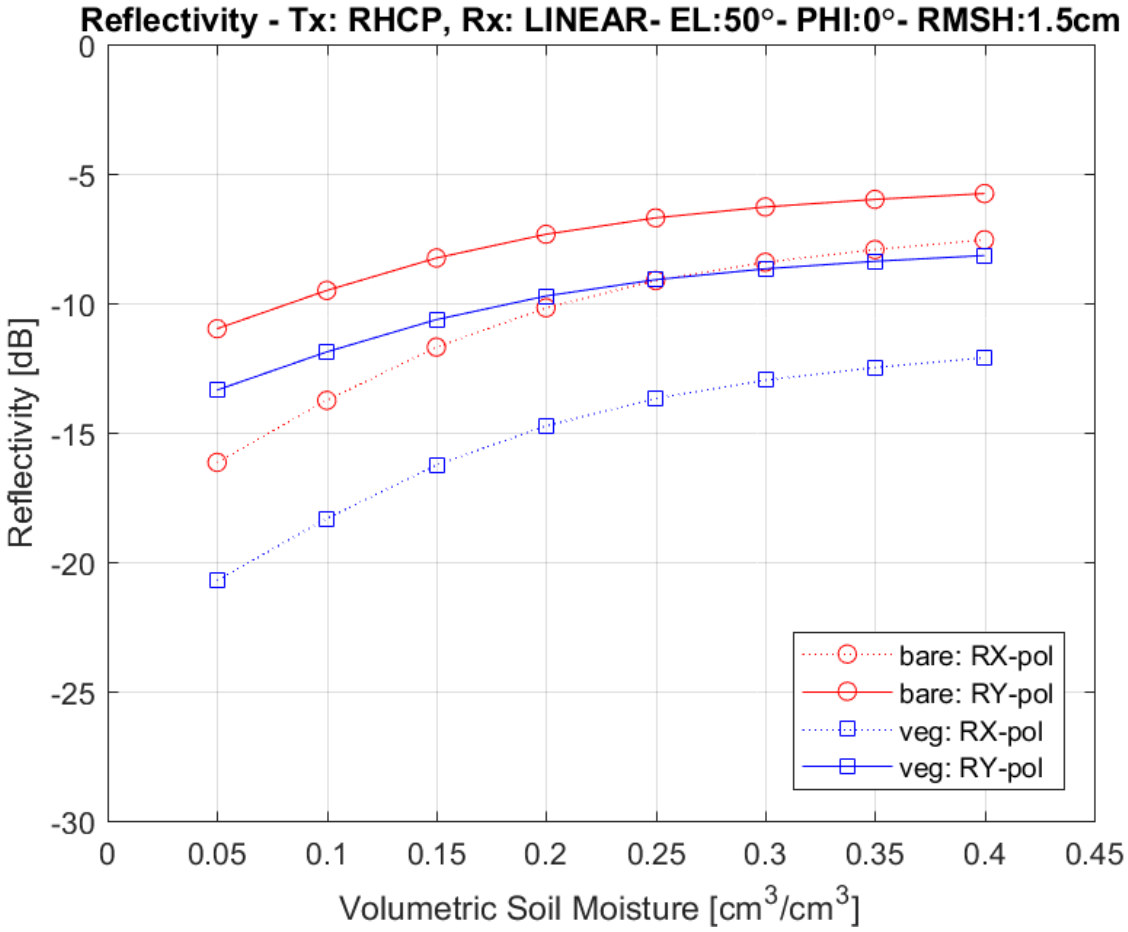

Figure 3.6: Reflectivity as a function of VSM for the default Forest inputs. Transmitter elevation angle is 50°, the surface RMSH roughness is 1.5 cm.

receiver is similar to that of the Forest defaults except that a right hand circular polarization (RHCP) antenna is assumed in this case. The vegetation layer inputs are prepared considering the real-world statistics of the corn stalks, leaves, and cobs. This *Agriculture* input configuration has been exploited in a corn simulation study, where GNSS-R responses to dynamic scene conditions are quantified [51]. This example is explored in greater detail within the SCoBi user manual [48]. The simulated reflectivity as a function of the transmitter elevation angle is depicted in Figure 3.7.

C. Soil

The default input for the Soil analysis allows to simulate a P-band SoOp configuration over bare-soil. The ground is considered to be single-layered; hence, this analysis covers a surface reflection scenario. This setup employs the same transmitter with the default *Agriculture* input and the same receiver with the *Forest* input. The main purpose of the default *Soil* inputs is to enable performing an L-band Times-series analysis over a bare-soil terrain. Further details about this example is included within the SCoBi user manual [48].

D. Root-zone

The default input for the *Root-zone* analysis is quite different than that for the previous analyses because it represents a multi-layered soil profile. A multi-layered bare-soil terrain is considered in this default input for the sake of simplicity. However, any Root-zone analysis may include vegetation cover as well. The default input uses the same transmitter satellite as the *Forest* analysis has. It considers a similar receiver as well, but the receiver

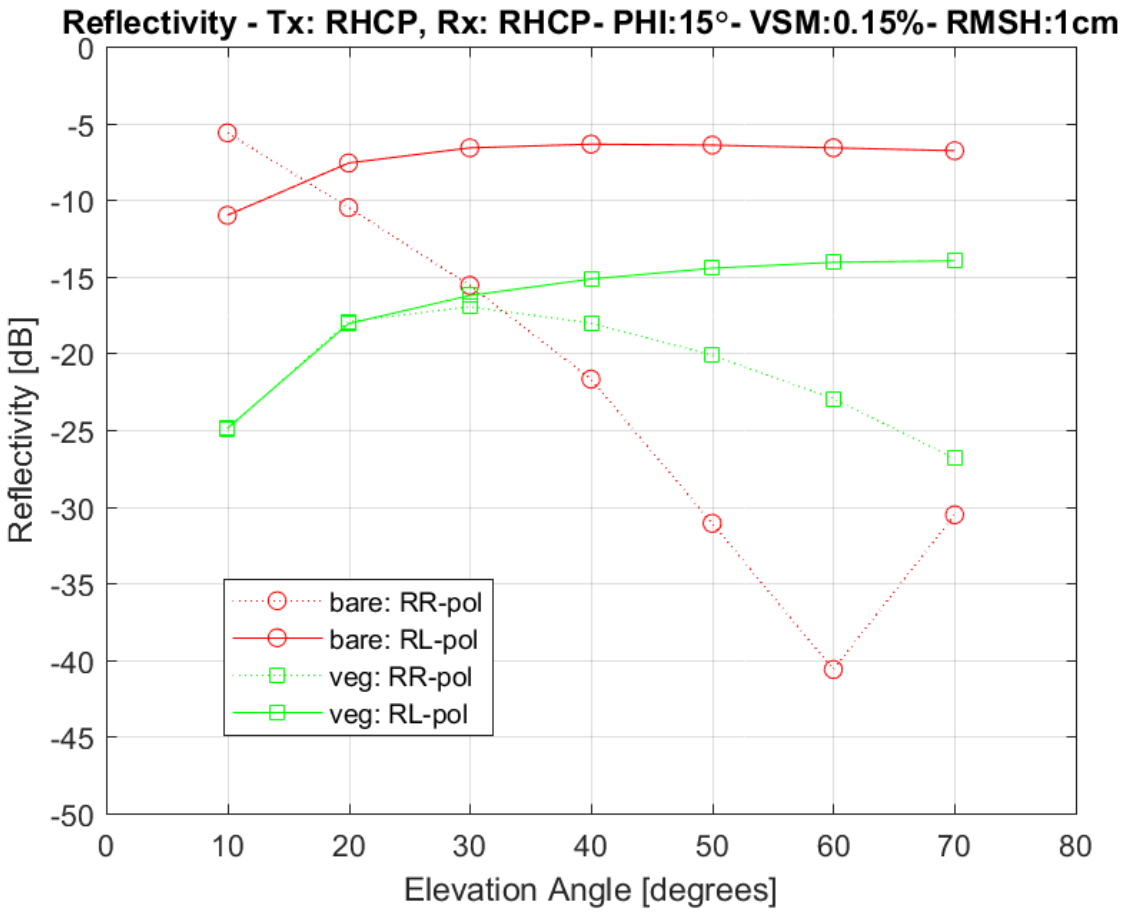

Figure 3.7: Reflectivity as a function of transmitter elevation for the default Agriculture inputs. Transmitter azimuth angle is 15°, VSM is 0.15 cm^3/cm^3, the surface RMSH roughness is 1 cm.

now has a set of fixed orientation angles. A detailed example is explored within the SCoBi user manual [48].

3.4 SCoBi Architecture

The SCoBi simulator framework has been developed by an iterative and incremental development process (including requirements analysis, design, implementation, testing, and deployment). For instance, the requirements analysis was mostly done during the creation of the SCoBi model [93] in several years, and then the simulator framework has been designed, implemented, and tested with increments. However, many iterations have been performed on the requirements, design, and implementation after results of the tests and findings of the studies [93, 51] accomplished by the preliminary SCoBi versions. Our tests have demonstrated the verification of the SCoBi simulator framework. In other words, tests show that the simulator meets the requirements of the SCoBi model. The validation of SCoBi is mature to some degree since the simulated results are compared to and shown to have agreements with several experimental studies in [93, 51] . The multi-layered ground simulations for bare-soil and vegetated terrains are currently being used for validation purposes [58] for the SNoOPI project [1]. CYGNSS-based reflectivity derivations [143] over homogenous and known terrain conditions (such as agricultural fields) are currently being compared to SCoBi-generated reflectivity values during our ongoing research [50]. Moreover, airborne and tower data collection campaigns are planned for further experimental validation in the near future. For this purpose, observatory data are currently being collected over several terrains by using in-house built drone-based receivers. The

maintenance of the SCoBi product will be handled by the authors for scheduled improvements (such as improving exception-handling mechanisms or adding new SoOp analysis types), user requests, and possible scientific collaborations.

The SCoBi source code has been implemented in the MATLAB R2017a environment; however, it is compatible with the versions above MATLAB R2015a (The oldest version which SCoBi was tested with) within MS Windows Operating System (Windows 10 64-bit). SCoBi does not require additional toolboxes or plugins of the MATLAB environment. MATLAB has been chosen for the development because of its common use among the researchers, efficient handling of the matrices, simple scripting features, and plotting capabilities. Both the structural and behavioral design models of the SCoBi software can be found in the Sparx Systems' Enterprise Architect design file that is shown in Figure 3.1. However, the SCoBi design details can be explained in this paper as follows:

A. Structural Design

The SCoBi architectural design is mainly achieved with the procedural programming (PP) principles that MATLAB intrinsically supports. However, object oriented programming (OOP) design and implementation principles are also utilized as needed for advanced design, data encapsulation, manipulation, code organization and readability, and maintenance purposes. Combining two design approaches is for the purpose of having the enumerated advantages of the OOP design while exploiting MATLAB's procedural scripting capabilities. For instance, the simulation engine (runSCoBi.m) is simply operated by a MATLAB procedure (function) implementation, whereas the dynamic and static system parameters are handled with the help of several classes with singleton pattern features. The

software packages within the source code are determined with respect to the relational hierarchy between each software entity (MATLAB functions or classes). Each package consists of several functions and/or classes. The UML (Unified Modeling Language) package diagram for the SCoBi source code ($/setminussource/setminuslib/setminus$) packages is shown in Figure 3.8. The *runSCoBi.m* function is directly under the *lib* package. It uses several packages to perform specific tasks in order to compute the model's output; for instance, the *gui* package is used to obtain the user inputs for simulations, *init* and *param* packages are used to initialize the simulation parameters by using the information (inputs) from the *gui* package, and the *main* package is used to perform every simulation iteration. The *main* package uses the *param* package to manipulate parameters-related tasks, the *bistatic* package to handle the bistatic geometry, the *ground* package to account for ground operations (dielectric calculation, and specular reflection), the *multilayer* and *vegetation* packages if involved in a simulation, and the *products* package to create and store simulation outputs. There are information flows from the *bistatic, ground, multilayer* (if included), and *vegetation* (if included) packages to the *products* package.

The SCoBi design file (created in the Enterprise Architect) also includes the UML class model of the software (*lib* package). In fact, class models are dedicated to the class instances in the OOP designs; however, this model is employed to depict the entire structural relations (usage, information flow, or inheritance) between the source code entities (MATLAB functions and classes). Although this is not a valid use of the class model, it can help the developers understand the general structure of SCoBi. Because the overall SCoBi class model is highly complicated, class models are also provided for the *runSCoBi.m* and

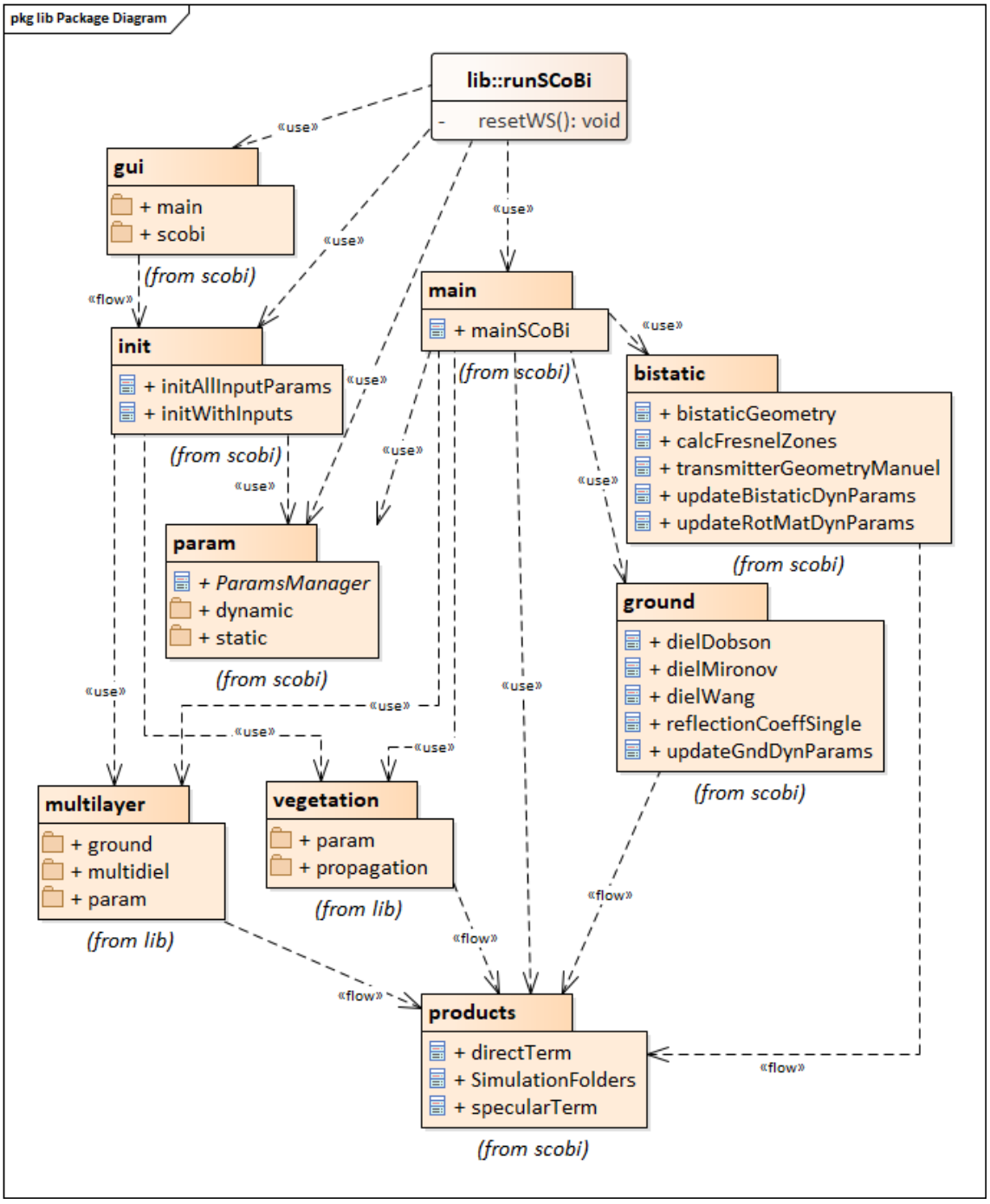

Figure 3.8: SCoBi ”lib” package diagram

mainSCoBi.m functions, which are the simulation engine and simulation iterator functions, respectively. These two models only show the dedicated function and its first-degree relations with the other source code entities.

B. Behavioral Design

UML behavioral diagrams such as sequence and activity diagrams are specialized to demonstrate the dynamic aspects of software programs. The UML activity diagrams are used to show the temporal flow of the SCoBi simulator. The SCoBi design file contains two activity diagrams for both the simulation engine and simulation iterator since these two deals with the overall flow. The SCoBi simulator framework always starts with running the *runSCoBi.m* function. The activities and the decisions that are performed within this function are shown in Figure 3.9. In summary, it gets the user inputs, initializes the parameters and the simulation by using these inputs with the help of input validation controls, and calls the simulation iterator (*mainSCoBi.m*) after writing the simulation reports.

The *runSCoBi.m* determines the required number of simulation iterations for a chosen simulation by using the parameters manager class. This is because the simulation mode, *Snapshot* or *Time-series*, in conjunction with the system and configuration inputs may change the number of total simulation iterations. Details of this phenomenon are described in both the user's manual [48] and the tutorial videos [9]. The simulation engine then runs the simulation iteration function (*mainSCoBi.m*) for the number of simulations.

The simulation iteration function, *mainSCoBi.m*, is always the same regardless of the analysis type. However, the analysis type and input selections affect the package usage and procedure calls within *mainSCoBi.m*. For instance, a vegetation analysis requires SCoBi

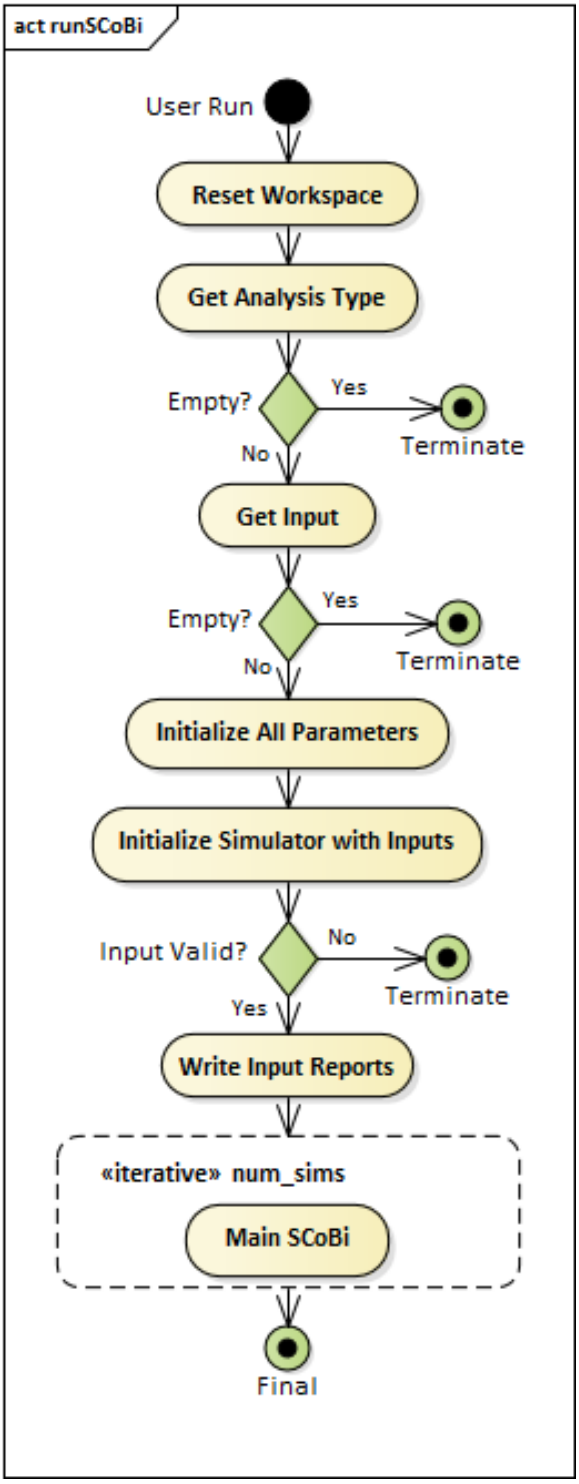

Figure 3.9: runSCoBi.m (Simulation engine) activity diagram

to use the software package, *vegetation*, and call several related procedures such as propagation calculation.

The flow of the *mainSCoBi.m* function is shown in a UML activity diagram in Figure 3.10. Since this function is called in every iteration of the simulation engine, it handles the updates and calculations related to the iterative steps. For instance, it starts with updating the dynamic parameters for the bistatic geometry and ground surface. These updates are needed in each iteration since the simulation inputs may contain changing transmitter orientation angles and land geophysical parameters such as VSM and RMSH. If the *Ground Structure* is *Multi-layered*, the multi-layer dielectric profiles are computed. If the Ground Cover is Vegetation, the vegetation propagation is calculated. Furthermore, the vegetation attenuation values are written into an Excel file if chosen by the user. The polarization rotation matrices for the local antenna coordinate systems are also required to be updated due to the changes in the bistatic geometry in each iteration. Finally, the current iteration gets ready to generate the SoOp deliverables (the direct and specular contributions) for the parameters in the iteration, and stores the iteration calculations incrementally to the simulation outputs.

3.5 Future Developments

The SCoBi simulator has been designed with the intention to foster a community of developers who may advance the modeling and simulation capabilities for SoOp over land applications. The framework is designed to simplify the implementation of new features

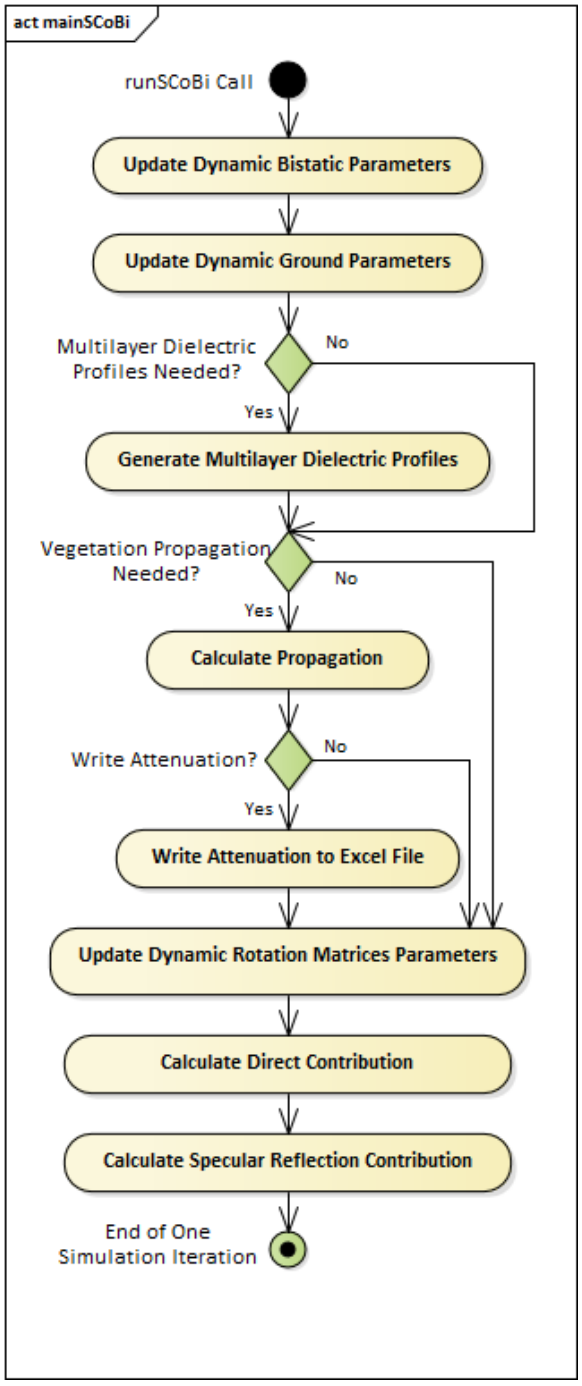

Figure 3.10: mainSCoBi (Simulation iteration function) activity diagram

for the simulator. A developer's manual is included within the SCoBi distribution package to encourage and assist developers to extend this SoOp simulator.

The potential future extensions to the SCoBi framework are two folds: improvements by its initial developers, or improvement by the community for specific needs. We, as the current SCoBi developers, aim to organize our future extensions in two directions: improving the structure of the SCoBi framework and adding new capabilities. Implementing advanced exception handling mechanisms can be considered as one of the major structural improvements to strengthen the current architecture, although the current SCoBi provides several handling techniques such as input manipulation (regulation) and use of management classes. Batch running without any GUI window can be another objective to make SCoBi more flexible to be used for fast analysis purposes.

The second class of improvements that we aim to obtain within a couple of future versions is to enlarge SCoBi's features. For instance, the visualization capabilities are as significant as the generating the simulated observables. We plan to release the plotting options for Time-series analysis over Multi-layered grounds as well as providing user friendly GUI windows for visualization purposes.

SCoBi can also be extended in response to Earth science community's needs. Such extensions can be made by us upon the requests, by collaborations, or by the future users' and developers' own initiatives. To illustrate, the diffuse scattering, which is already handled by the SCoBi model [93], might be added to the simulator if any specific need arises, although we believe diffuse contribution is far from dominating SoOp land observables. In addition, the SCoBi forward model possesses many capabilities that are not yet utilized by

the SCoBi simulator, although the current distribution contains many novel tools and features. For instance, the model support the use of new antenna pattern methods; thus, a new antenna pattern generation such as ”cosine to power n” can be added to the framework. Similarly, geo-synchronous transmitters, a new ground dielectric model, or a new plotting function can be incorporated into the framework. For those who try to extend SCoBi alone, the SCoBi developer's manual and UML designs can help much. A major extension could be the implementation of one of the unimplemented analysis types. For example, snow cover on the ground, topographic relief to the current scattering calculations, permafrost analysis within the root-zone, or wetland analysis can be introduced to the model.

3.6 Conclusion

A new, fully polarimetric, coherent SoOp simulator has been introduced for the Earth science community. It has unprecedented capabilities such as supporting any combination of polarizations, taking into account the significant antenna characteristics (polarization crosstalk and mismatch, different antenna pattern methods), handling vegetation and multi-layered soil analysis at the same time, supporting multiple models for soil dielectric calculations, and enabling interferometric analysis by providing the complex field quantities. It simulates the direct and specular reflection contributions on the bistatically received signals. It is released with a set of built-in plotting functions. Furthermore, it has a user-friendly GUI; it is free and open-source with the user's manual [48] and developer's manual [46] together with a set of tutorial videos [9]. The SCoBi simulator framework is aimed at creating an environment where both lifelong researchers and newcomers to SoOp re-

flectometry can investigate new methods for SoOp Earth science applications. It can be exploited by researchers, scientists, and even end-users with little-or-no electromagnetic background. SCoBi has the potential **(i)** to allow its users to study new SoOp techniques, **(ii)** to make analyzing the SoOp methods in detail possible, **(iii)** to help determine optimal configurations for directed use, and **(iv)** to enable generating, visualizing, and analyzing large amount of simulated test data.

CHAPTER 4

RESPONSE OF GNSS-R ON DYNAMIC VEGETATED TERRAIN CONDITIONS

This chapter is inherited from the following publication:

O. Eroglu, M. Kurum, and J. Ball, "Reponse of GNSS-R on dynamic vegetated terrain conditions," *IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing*, vol. 12, no. 5, 2019, pp. 1599-1611.

Significant author contributions to the above publication were as follows: Conceptualization and Methodology by **Eroglu** and Kurum; Investigation and Validation by **Eroglu**, Kurum, and Ball; Writing by **Eroglu**; Review and Editing by **Eroglu**, Kurum, and Ball; Visualization by **Eroglu**; Supervision by Kurum; Project Administration by **Eroglu** and Kurum; Funding acquisition by **Eroglu** and Kurum.

Global Navigation Satellite System Reflectometry (GNSS-R) has the potential to offer a cost-effective solution for global land observations. In this study, we aim to understand GNSS-R sensitivity to changing land geophysical parameters. For this objective, we performed simulations of a ground-based receiver using a recently developed coherent bistatic vegetation scattering model (SCoBi-Veg) to detect GNSS-R signatures under varying soil moisture (SM), vegetation water content (VWC), and surface roughness during a full corn growing season. We modeled different corn growth stages by using *in situ* measurement

data. We analyzed the simulated reflectivity and received power values based on the aforementioned variable input parameters. This study demonstrates that specular reflections dominate the diffusely scattered contribution in case of moderate roughness regardless of the corn field row structure or the polarization. Significant correlations between VWC and cross-polarized reflectivity values are also shown. Furthermore, the study quantifies the effects of SM and surface roughness on GNSS-R deliverables.

4.1 Introduction

Soil moisture (SM) is a key land geophysical parameter on the water, energy, and carbon fluxes through the ground-air boundary of the earth because it is a major component governing the high energy-demanding evapotranspiration processes. Understanding the SM with high spatial and temporal resolutions at global scales can help improve applications such as crop yield, weather, and natural hazards (e.g. flood, landslide, and drought) forecast [41, 103, 30].

Global SM sensing is a difficult, complex problem due to challenges arising from sensor technologies, observation geometry, atmospheric losses, and spatial SM variations. Microwave remote sensing, particularly at L-band, has been used for this purpose for decades because of its sensitivity to SM and other geophysical parameters [168]. Conventional microwave methods use passive or active mono-static instruments such as radiometers or radars, respectively. For example, NASA's Soil Moisture Active Passive (SMAP) [42], ESA's Soil Moisture Ocean Salinity (SMOS) [85], and Japan's Phased Array L-band Syn-

thetic Aperture Radar (PALSAR) [137] are current space missions with L-band mono-static radars and/or radiometers that monitor geophysical parameters.

Key limitations of such traditional remote sensing methods are problems with their spatiotemporal sampling, which are often characterized by either coarse resolution with frequent revisit times, or high spatial but low temporal resolution. For instance, the spatial resolutions are 25-50 km for radiometers and tens of meters to few km for mono-static radars while the temporal revisit times vary from few days for radiometers to few weeks for radars. This hampers the use of such techniques to effectively observe SM for agriculture and hydrometeorology applications that often require both high spatial and high temporal resolutions [10]. The Signals of Opportunity (SoOp) technique, on the other hand, has recently gained an increasing attention in Earth science research community as a promising approach to improve spatio-temporal resolutions for microwave remote sensing of SM. It is based on reception and assessment of the reflected signals from the Earth's surface in a multi-static scheme, which are already emitted by existing illuminators such as navigation and communication satellites [178]. In particular, Global Navigation Satellite System (GNSS) Reflectometry (GNSS-R) has been widely studied as an application domain of SoOp. The large number of satellites in any GNSS constellation (such as GPS, GLONASS, Galileo, and BeiDou) in the medium Earth orbit allows for frequent revisit times [70]. Several small-form factor GNSS-R satellites can be launched in a GNSS-R constellation as well, which also improves the revisit times. Moreover, the spatial resolution for land can fall under even one km in theory since the coherent (specular) reflection through the first Fresnel zone is expected to be dominant [82]. Thus, GNSS-R has the

potential to provide global coverage with high spatiotemporal resolution by using small SWaP-C (Size Weight and Power - Cost) receivers in small satellites.

Several recent studies examined GNSS-R suitability for land remote sensing. Most of them are field experiments with towers or airborne instruments. A reasonable number of these studies measure the interference between the direct and the reflected signals by means of a ground-based, horizontal-looking, geodetic GNSS receiver. Examples of experimental studies that employ this method to retrieve SM, vegetation height, and surface topography over bare or vegetated terrains are reported in [82, 133, 134, 132, 4, 97, 25]. Some other ground-based polarimetric measurements were also performed [39, 40]. There exist a number of airborne experiments that exploit specially designed GNSS receivers to use polarimetric measurements for SM and biomass sensing [82, 114, 12, 65].

Several researchers focused on the use of the GNSS-R approach from space such as the qualitative analysis of the UK TechDemoSat-1 (TDS-1) data over land [13, 23, 118]. The major contribution of this group is to evaluate the feasibility of TDS-1 observations for land applications. Generally, such qualitative works compile comparisons and possible correlations between TDS-1 land returns and other satellite missions that are dedicated to particular retrieval tasks. For example, Camps et al. [13] demonstrate TDS-1's sensitivity to a wide range of SM over terrains with various vegetation covers. They perform a sensitivity analysis by comparing TDS-1 data to the SM data of Soil Moisture Ocean Salinity (SMOS) and normalized difference vegetation index (NDVI) data of Moderate Resolution Imaging Spectro-radiometer (MODIS) missions.

Very few studies utilize scattering models for detailed analysis of factors influencing the received signal in bistatic configuration. Such studies are mainly based on either radiative transfer theory [106, 52, 66, 130, 174], or analytical wave theory in conjunction with distorted Born approximation [155, 92, 93]. The former group consists of models that combine the coherent specular reflectivity with the incoherent scattering using the energy conservation principle, while the latter group includes coherent models that provide coherent sum of complex scattered fields. These bistatic models have been applied to diverse terrains such as bare soil, crop fields, and tree canopies under various combinations of linear and circular polarizations and microwave configurations ranging from P to S-bands. In fact, such a wide applicability is the major strength of the model and simulation studies; they can enable Earth science community to explore bistatic configurations and methods.

The main issue for application of GNSS-R to spaceborne instruments stems form the incapability of experimental and comparative studies to quantify the joint impact of distinct parameters, and the insufficient modeling effort that can accurately describe the effect of SM and other parameters. Besides, such a shortcoming often causes some puzzling conclusions through the interpretations of the results. For instance, Camps et al. [13] observed diverse SNR and reflectivity trends for bare and sparsely vegetated terrains with very similar surface characteristics. This may be due to the fact that high topographic relief may have led to a dramatic increase in the incoherent scattering, whereas the coherent reflection dominates in other land observations. Furthermore, it may have been caused by spatial SM variations as well. Chew et al. [23] found that the direct proportion between the reflectivity and the SM gets satisfied with a sufficiently large dielectric constant for some cases. All in

all, such puzzling or qualitative findings indicate the further work needed to quantify the conjunct impacts of different geophysical parameters.

The key contributions of this study is to model a realistic corn crop for accurately representing the vegetation effects, to demonstrate GNSS-R signatures for eliminating the previous puzzling findings, and to help develop generic analytical models for geophysical land parameter retrieval. Therefore, this paper aims to (1) quantitatively compare specular and diffuse contributions by analyzing the co-polarized and cross-polarized received power for varying vegetation and ground conditions, (2) analyze correlations between the co-polarized and cross-polarized GNSS-R signals and the dynamic parameters (such as SM, vegetation water content (VWC), and surface roughness). In order to achieve these objectives, the study employs our recently developed coherent bistatic vegetation scattering model titled **S**ignals of Opportunity **Co**herent **Bi**static Scattering Model for **Veg**etated Terrains (SCoBi-Veg) [93], and its simulator [45].

The next sections are organized as follows: Section II describes the SCoBi-Veg model that is employed in the simulations. It then provides the details of the simulation setup and the virtual corn modeling. Section III illustrates the results of the simulations and presents a comprehensive discussion about significant findings. Section IV concludes the major aspects and the potential future of the study.

4.2 SCoBi-Veg Model and Simulation Setup

The SCoBi-Veg model and the simulator predicts polarimetric bistatic reflections over vegetated or bare terrains, where the ground is assumed to be flat with a smooth to moderate

surface roughness level [93]. It is specialized to the most common SoOp configuration, which consists of a transmitter far-away from the Earth surface and a receiver at a low altitude. Figure 4.1 depicts such arrangement over a row-structured crop field. The Fresnel zones that are discriminated by multiples of half-wavelength (unique phase differences) and the receiver footprint are also shown. The East-North-Up (ENU) coordinate system is used as a reference frame to relate the transmitter, receiver, and local coordinates to each other. In the figure, the point S represents the specular reflection point, where the incident and reflected waves make the same angle of θ_s with the surface normal. The azimuth orientation is represented by the angle ϕ_s, which is defined between the ground range and the local north direction. The quantities h_r and h_t represent the receiver and transmitter altitudes, respectively. The range term r_d denotes the line-of-sight distance between the transmitter and the receiver, r_{st} is the distance between the specular reflection point and the transmitter, and r_{sr} stands for the distance between the specular reflection point and the receiver.

SCoBi-Veg applies analytical wave theory in conjunction with the distorted Born approximation to describe vegetation propagation and scattering. The surface roughness under the vegetation is smooth and follows Kirchhoff's approximation with a Gaussian height distribution, so no incoherent scattering from ground is considered. As a typical forward model, SCoBi-Veg demands several input parameters such as transmitter and receiver specifications and orientations, vegetation characteristics, ground soil texture, roughness, and soil moisture. The transmitter and the receiver effects such as polarization crosstalk and mismatch, and orientation are explicitly included through normalized radiation pattern ma-

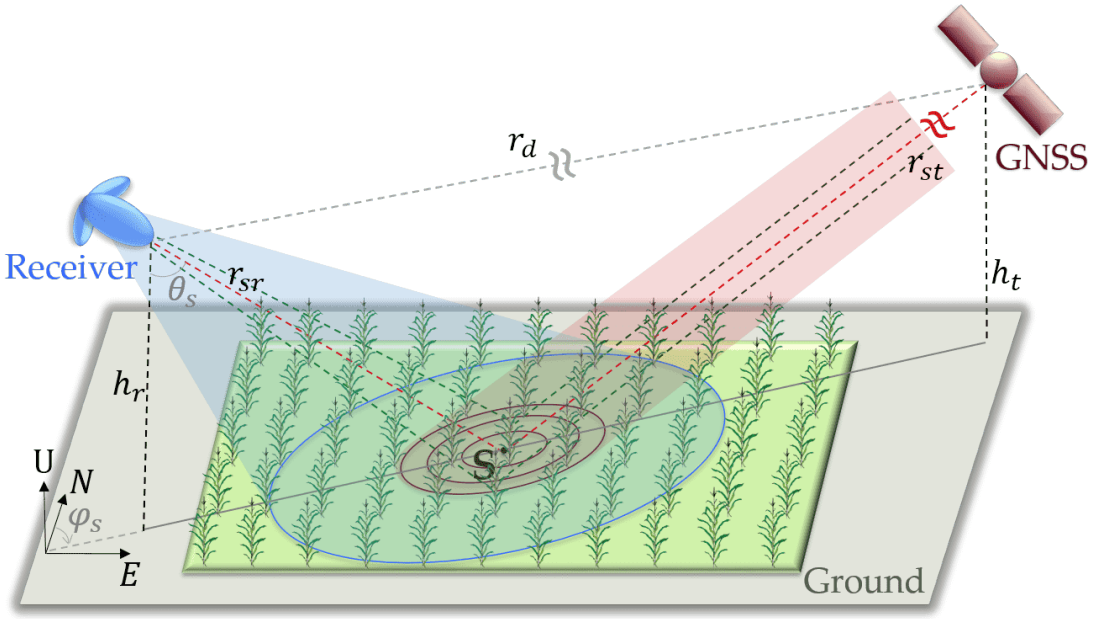

Figure 4.1: Bistatic configuration of the simulations. Transmitter is a GNSS satellite. Receiver is a ground-based, dual-pol antenna always pointing toward the specular reflection point. Vegetation cover is a row-structured corn field.

trices and the polarization basis rotation matrices. The model calculates received specular (coherent), diffuse (incoherent), and direct contributions independently. In addition to the phase information retained for interferometric analysis, the complex field quantities are also converted into the modified Stokes intensity vectors, given by:

$$\underline{P}_m^{coh} = \left\{ \frac{|\ K\ |^2}{(r_{st} + r_{sr})^2} \right\} \underline{\Gamma}_s \tag{4.1}$$

$$\underline{P}_m^{inc} = \left\{ \frac{|\ K\ |^2\ A_s}{4\pi r_{st}^2 r_{sr}^2} \right\} \underline{\sigma}_e^0 \tag{4.2}$$

$$\underline{P}_d = \left\{ \frac{|\ K\ |^2}{r_d^2} \right\} \underline{D} \tag{4.3}$$

where

$$\mid K \mid^2 = \left(\frac{\lambda_0}{4\pi}\right)^2 G_{0r} EIRP \tag{4.4}$$

From (1) through (3), each underlined term represents a four-by-one modified Stokes vector to describe the wave intensity and polarization. The vectors $\underline{D}$, $\underline{\Gamma}_s$, and $\underline{\sigma}_e^0$ denote direct signal, specular reflectivity, and effective normalized bistatic radar cross section (NBRCS), respectively. The quantity $\mid K \mid^2$ is a constant that involves free space wavelength (λ_0), the maximum gain of the receiver antenna (G_{0r}) and the Equivalent Isotropically Radiated Power ($EIRP$) of the transmitter. The quantity A_s is the contributing surface area, where the diffuse scattering originates.

The specular reflectivity $\underline{\Gamma}_s$ involves the wave attenuation as the wave propagates down from the top of the vegetation to the ground, reflection from the surface specularly, and again attenuation from the ground to the top of the vegetation. The surface reflection is assumed to be coherent where smooth-ground Fresnel reflection is used. The reflectivity is essentially attenuated twice by the vegetation. The attenuation and phase change of the coherent wave, propagating in the equivalent medium, is found by calculating the mean field within the medium.

The "effective" NBRCS is acquired through a Monte Carlo scheme that generates a sufficient number of realizations for canonical scatterers within the vegetation canopy. It includes the spreading loss, and the path-dependent phase terms for different diffuse mechanisms in the various multipath directions for all the scattering particles. An average effective NBRCS is obtained by 20 realizations of the vegetation layer through Monte

Carlo simulations for the simulations described next section. The term effective" in front of NBRCS comes from the fact that the antenna characteristics as well as the statistical and physical properties of the terrain are blended in NBRCS, since the antenna radiation pattern projected on the surface is not uniformly distributed in phase, amplitude, or polarization for ground-based systems.

Finally, the direct signal $\underline{D}$ represents the signal received through direct line of sight and is included in this section for the sake of completeness even though the direct signal is not used in the analysis here. Expansion of the terms in the equations from (1) to (4) and further details can be found in [93].

4.2.1 Simulation Setup

This study aims to simulate GNSS-R response to a row structured crop field under dynamic scene conditions and to observe its sensitivity to several geophysical and system parameters. To generate large synthetic data, snapshots of combinations of a set of these dynamic parameters were simulated for a typical GNSS-R bistatic geometry as illustrated in Figure 4.1. The transmitter is a GPS satellite operating with right hand circular polarization (RHCP) at 1575.42 MHz (L1 C/A-band) with a fixed EIRP of 27 decibels (dB). The receiver is a dual-polarized (RHCP and LHCPleft hand circular polarization) antenna system situated at 20 m above ground (a typical tower height). The receiver antenna is modelled as a circularly symmetrical generalized Gaussian pattern with a half-power beamwidth of 30°, side-lobe level of 30 dB, and cross polarization levels of 15 dB, 25 dB, and 40 dB. The coherent and incoherent processing receiver gain can also be added to the received

power in (1)-(3), but it will not change the sensitivity other than introducing an offset to the power values, provided that the thermal noise is not considered. The main beam axis of the receiving antenna is chosen to always face at the specular reflection point, which can be seen in Figure 4.1. This is for the purpose of directly analyzing the GNSS-R's angular sensitivity for varying incidence angles by minimizing the antenna pattern effects and polarization mismatch, which might significantly dominate the ground-based observations. As a result, receiver's zenith observation angle is set equal to the transmitter incidence angle (90° - transmitter elevation angle). If the receiver was a nadir-looking antenna, it would require a well-designed omnidirectional pattern to obtain similar results to the present configuration. At the chosen altitude (due to its relatively close proximity to the ground), the receiver antenna footprint limits the contributing area. In other words, diffuse scattering signals coming from the one GNSS chip length (around 300 m) are spatially filtered by the antenna footprint, which falls into approximately the first ten Fresnel zones for the given configuration. Hence, the diffuse scattering power is calculated over the entire footprint (by using ten Fresnel zones); however, effective NBRCS is only considered from the first Fresnel zone to minimize the effect of the system parameters.

The dynamic geophysical and system parameters are the corn growth stages, the zenith observation (θ_s) and azimuth observation (ϕ_s) angles, volumetric SM (VSM), and root mean square height (RMSH) roughness of the surface. Modeling of the corn growth season will be detailed in the next subsection. The zenith observation angle ranges from 10° to 70° by 10° increments. The azimuth angle between the antenna pointing and the plant rows is evaluated for 0°, 30°, 45°, 60°, or 90°. VSM is varied in the interval of [0.05, 0.40]

cm^3/cm^3 by 0.05 cm^3/cm^3 increments. The crop ground is assumed to be flat and to have a typical surface roughness for agriculture, which is modelled uniformly over the ground by three different RMSH values of 0.5, 1.0, and 1.5 cm. For every combination of the dynamic parameter values in conjunction with separate corn growth stages, SCoBi-Veg generates a new database of simulated GNSS-R observations that consist of the direct, specular, and diffuse contributions.

4.2.2 Corn Field Modeling

The corn growth stage is the primary dynamic input of the simulating system since it comprises vegetation geometric structure and statistics as well as VWC throughout the growing season. Physiologically, corn development consists of two major phases: vegetative and reproductive, as shown in Figure 4.2. The vegetative phase is the period of rapid growth without cob generation, and it includes VE (Emergence), V1 (one leaf) to Vn (n leaves), and VT (Tassel). The reproductive phase is the later period with almost constant plant constituent dimensions, and it involves from R1 (Silking) to R6 (Maturity). Briefly, the cobs develop from R1 to R4. R5 (Dent) is the stage where the kernels are mostly dented at a moisture level around 55%. R6 is the stage with the lowest overall moisture level, where the stalks might still have a moisture content level under their yellowish shell, the leaves are totally dry, and the cobs have a moisture level around 35% [27].

In this study, *in situ* measurement data from the ComRAD (Combined Radar Radiometer) 2012 field campaign in Maryland [123] is utilized to account for the temporal evolution of the corn crop. It is a comprehensive study of detailed corn field observations, which in-

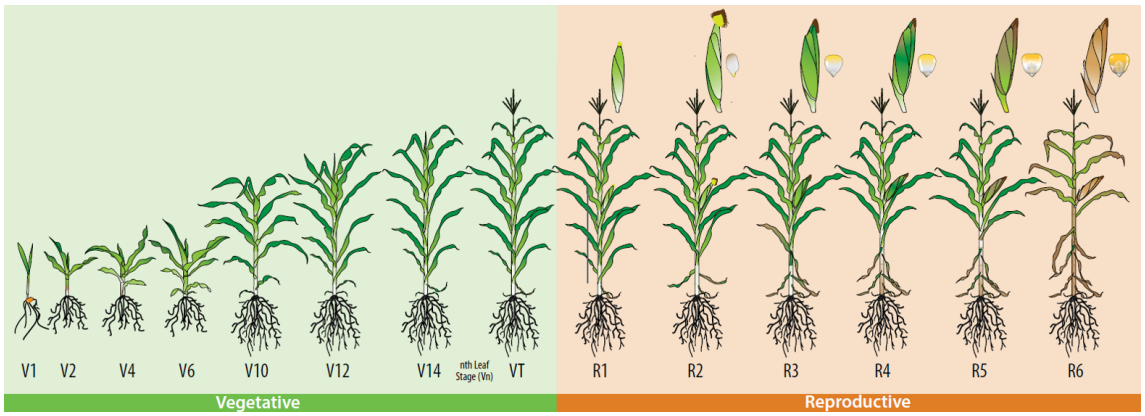

Figure 4.2: Corn growth stages for Vegetative and Reproductive Phases [27]. The vegetative stages are named by the number of leaves, whereas the reproductive stages are named regarding the cob development.

clude the measurements of plant architecture (dimensions and orientations of stalk, leaves, and cobs), density, VWC and dielectric constant of vegetation constituents. Canopy measurements except the dielectric constant data were collected roughly once every week during the entire growing season, some instances of which are shown in Figure 4.3a, b, c. In addition, Figure 4.3d shows canonical shapes that represent vegetation constituents in this study such as thin dielectric discs for leaves [104, 105] and dielectric cylinders for the stalks and cobs [147, 79]. The dielectric constants of the corn stalks, leaves, and cobs were measured at only five different dates during the season (13 July, 1 and 23 August, 8 and 20 September). To be more precise about the *in situ* vegetation data set, it includes the height, bottom and top node diameters, and multiple node/internode dielectric constants of the stalk; the length, width, and dielectric constant of the leaf; and the length, diameter, and multi-point (outer, inner, bottom, center) dielectric constants of the cob. The soil is characterized by a soil texture with a sand fraction of 0.60, clay fraction of 0.16, and bulk

density of 1.25 g/cm^3. It should be also noted here that the year of the experiment was very dry, which affected the measurements. However, it is not a critical constraint on analyzing the relative sensitivity of the GNSS-R to varying parameters.

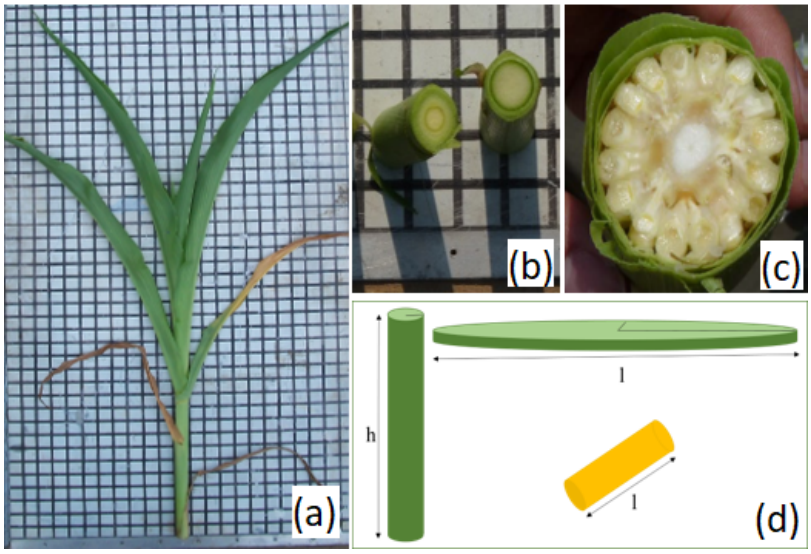

Figure 4.3: *In situ* measurements and virtual constituent representations. (a) Corn dimension and orientation (b) Stalk (c) Cob (d) Cylinder and elliptic disk to represent stalk and cob, and leaf, respectively.

The corn growth stages are divided into five different groups to represent the variations in size and moisture content throughout the season by separating the measurement data into five intervals. This is done with respect to VWC amount, constituent dimensions, and cob existence in each measurement period. The two vegetative stage groups are represented by V1-V9 and V10-VT while the three reproductive stage groups are denoted by R1-R4, R5, and R6. The V1-V9 stage represents the rapid emergence of the plant in its infancy period whereas the V10-VT stage accounts for the next rapid growth until the reproductive phase. The R1-R4 phase corresponds to the highest VWC level and largest constituent

dimensions (including the cob) within the season while the R5 and R6 periods represent the physiological stages with the same names and their corresponding conditions as described before.

Figure 4.4 provides the VWC and dielectric sampling dates, and measured VWC values for stalks, leaves, and cobs and the growth stage separations with different colors. As can be seen from the figure, V1-V9 spans from the emergence to June 26, V10-VT is modelled from the data between June 26 and July 17, R1-R4 covers the middle of the season, R5 is in the interval that starts with August 15 and ends with September 19, and R6 stands in the senescence period. Measured dielectric data set is available for all growth stages except V1-V9, whose dielectric constant is assumed to be equal to that of the V10-VT due to its close phenomenology. Figure 4.4 also shows that stalk and leaf VWC levels increase as the corn grows until the mid-season (R1-R4), then decrease gradually to the maturity (R6). The cobs emerge in the last week of July 20 and then follow similar growth trend. All in all, R1-R4 has the highest overall VWC whereas V1-V9 and R6 represent the lowest levels. V10-VT and R5 have similar VWC amounts; however, it should be kept in mind that there are cobs in R5 stage that can be an important source of diffuse scattering.

The corn canopy, both individual plant architecture and the entire crop-field structure, is realistically modelled as a detailed virtual 3-D row-crop field for the sake of accurately simulating the scattering. Every corn plant is a realistic conjoint body with a stalk that is vertical and fixed on the ground, and corresponding leaves and cobs that are attached to that stalk. Cobs are inserted to the stalk between one and three fourths of the stalk height, while leaves are inserted above one fifth of the height through the top of the stalk. To cre-

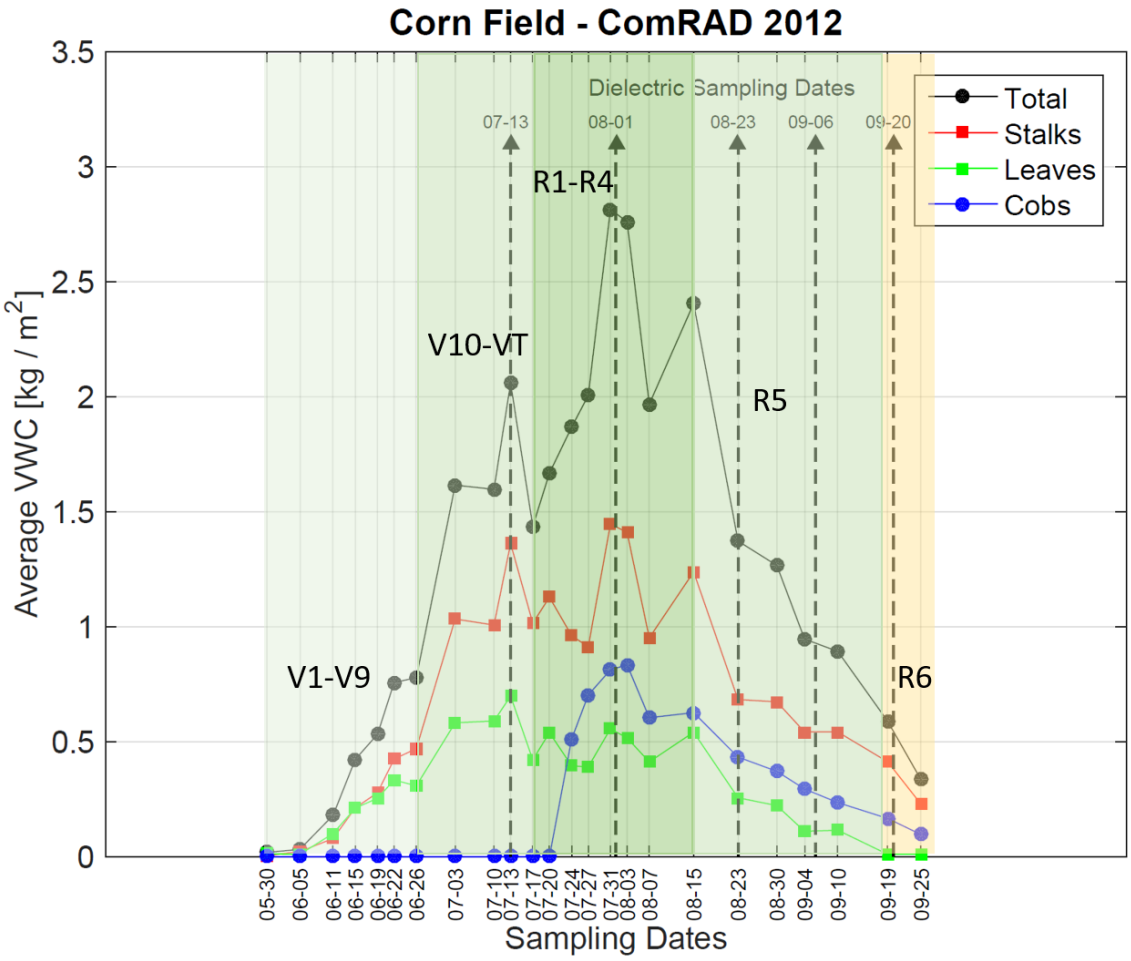

Figure 4.4: VWC measurement dates and values, dielectric measurement dates, and growth stage separation.

ate every individual constituent, *in situ* measurements (dimensions, orientation angles, and dielectric constant) within the growth stage of interest are utilized by fitting distinct probability distributions to those measurements to generate random variables that represent that growth stage. In other words, the orientation and dimensions of the stalk, leaves, and cobs in every individual corn plant are generated by a random process regarding the obtained probability distributions from the field data. The dielectric constants to the constituents are assigned by using the corresponding average dielectric measurement with a five percent standard deviation to accommodate possible variations within each stage. Due to the use of probability distribution functions, every corn plant becomes unique but represents the characteristics of its growth stage. In addition, all of the constituents are involved in the attenuation calculations, whereas leaves are neglected for the diffuse scattering calculations because the leaf thickness is much smaller than the wavelength.

Figure 4.5 shows the average measured dimensions and their standard deviations for the stalks, leaves, and cobs for each growth stage. The stalk height, leaf length and width, and cob length exhibit an apparently similar trend to the overall VWC levels, where they have lower values in the emergence and maturity stages and higher values through the mid-season. Hence, V10-VT and R5 generally have medium-scaled values between the season extrema. There exist a few unexpected dimensional behaviors as well. For instance, the stalk diameter slightly decreases during R1-R4 compared to the previous stages. Moreover, the cob diameter gradually increases throughout the entire season. Such behaviors seem abnormal for the physiological development of corn and are interpreted as the experimental

biases. However, in general, there is a consistent growth until R1-R4, then there is a shrinkage through the senescence stage.

The vegetation layer consists of rows, as visualized in Figure 4.1. The spacing between adjacent rows and corn seeds are 0.75 m and 0.20 m, respectively. Moreover, seed positions are generated by means of a quasi-periodic approach, where they may fluctuate in a small circle with a radius of 1.5 cm [164]. Figure 4.6 depicts the row structure and the quasi-periodic seed distribution for a small portion of the field. This setting results in a realistic row-crop with a plant density of around 6.7 stalks/m^2. In order to acquire the entire corn field, a sufficiently large square-shaped region that covers all ten Fresnel zones, which is enough for the entire antenna footprint, is first calculated. Then, the seed positions in this region are included in the simulation if they fall into any of the Fresnel zones. Depending on transmitter antenna incidence angles (from 10° to 70°), field size and plant numbers in the field are varied. In addition, the various azimuth angles (0°, 30°, 45°, 60°, and 90°) between the antenna orientation and the plant row are tested to inspect the effect of relative row orientations on GNSS-R response.

It is worth noting here that *in situ* VWC data are not directly used to model the corn vegetation in this study; instead, those are only exploited for the analysis of the simulation results, which is given in the next section. Hence, the dielectric characteristics of the vegetation layer is only represented by the *in situ* dielectric constants whose average real parts are shown in Figure 4.7 for each stage. The figure indicates that even though leaf and cob dielectric constants get very small through the senescence, stalk dielectric measurements tend to stay at a remarkably high level throughout the season. In order to explain the ef-

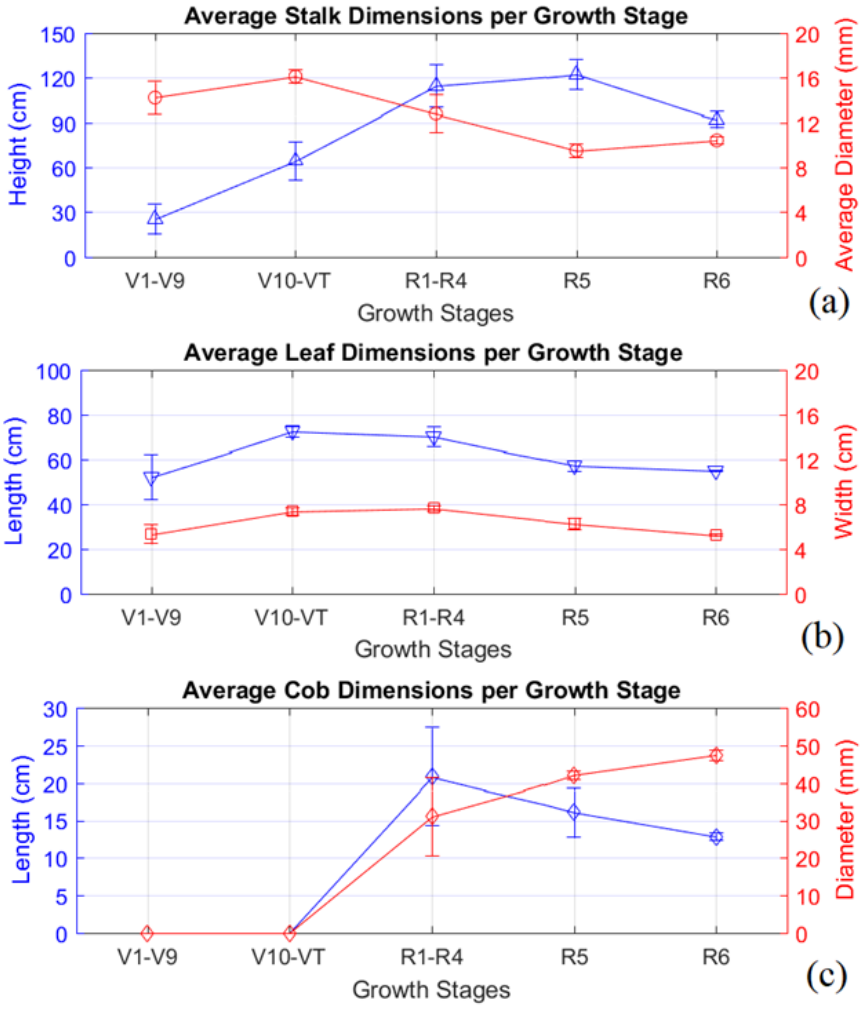

Figure 4.5: Average dimensions for varying growth stages. Error bars indicate the standard deviations. (a) Stalks, (b) Leaves, (c) Cobs.

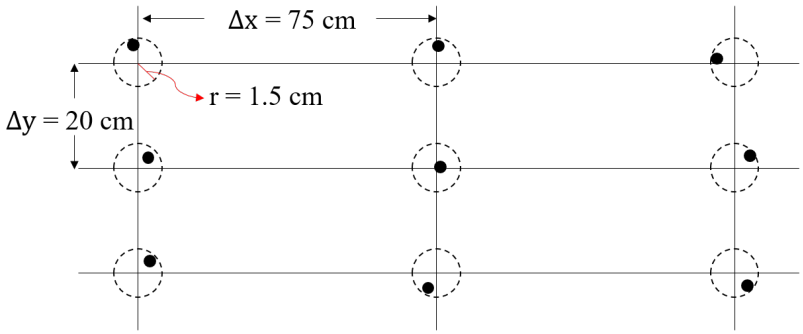

Figure 4.6: Quasi-periodic seed distribution within the row structure. Δx spacing between rows, Δy: spacing between adjacent seeds in a row, r: maximum radius of fluctuation circle.

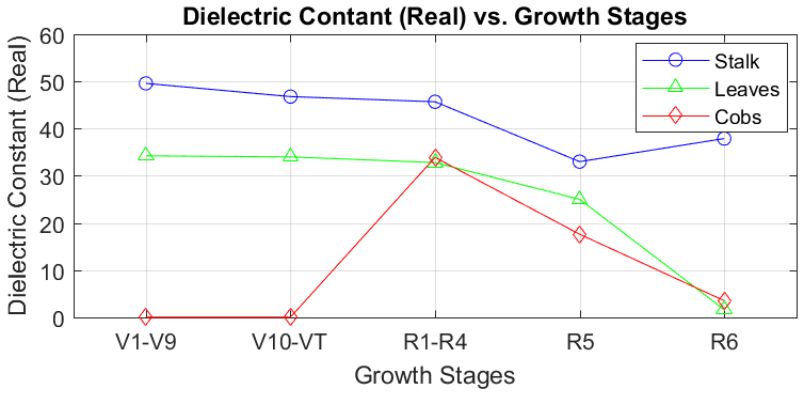

Figure 4.7: Average dielectric constant (real part) of the corn plant constituents.

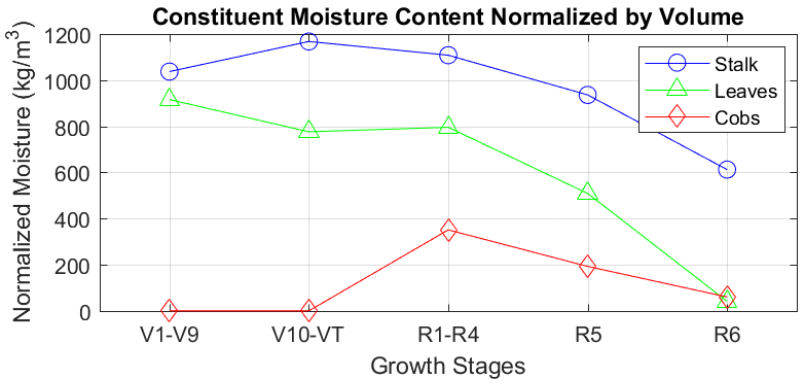

Figure 4.8: Normalized moisture content of the corn plant constituents by their volume per growth stages.

fect of moisture content on the measured dielectric constant values, the VWC of stalks, leaves, and cobs are normalized by the volumes of these constituents in each growth stage. The volumetic moisture contents (VWC per volume) are plotted for each constituents and stages in Figure 4.8. It is apparent from the figure that the leaves have high volumetric water content in early and mid stages, and then face a dramatic dehydration through the end of the season. The cobs show a rapid raise in moisture content until the R1-R4, then they dry out through the maturity, as also known from [27]. The figure also indicates that the stalks tend to save their effective moisture content even if they lose some amount since the variations in the dimensions appear to follow a parallel trend with the absolute moisture content. In other words, it can be seen by a comparison between Figure 4.5a and Figure 4.8 that expansions and shrinkages in the stalk dimensions make the normalized moisture content have smooth variations through the season. This phenomenon also seems consistent with the observations in the *in situ* field experiments, where mature stalks still have reasonable amount of moisture content despite their outer shells seem shrunk and entirely dry at the time of senescence. Such an observation on the normalized moisture content of the stalks is significant because it accounts for the dielectric constant of these scatterers over time. In other words, relative change of the stalk dimensions and moisture content avoids a drastic reduction in the volumetric moisture content through the growth season, which consequently prevents dielectric permittivity from severe decline. On the other hand, leaf and cob experience severe dehydration through the senescence, which leads to a lowering of normalized moisture content, and it well explains the low cob and leaf dielectric constants during R6.

4.3 Results and Discussion

In GNSS-R land applications, it has been often assumed that the coherent term is the dominant contributor to the received signal [82]. Kurum et al. [93] have recently provided a comprehensive discussion on the relative significance of the altitude dependence on the coherent and incoherent terms for low-altitude platforms. As can be seen in Eq. (1) and Eq. (2), the specular reflectivity is multiplied by $1/\left(r_{st}+r_{sr}\right)^2$ whereas the effective NBRCS is multiplied by $1/\left(4\pi r_{st}^2 r_{sr}^2\right)$. The multiplication in the diffuse term prevents it from having comparable levels against the specular contribution unless the surface is very rough or topographic relief exists. Also, note that the illuminated area of the incoherent return increases with height partially compensating the effect of the ranges, up to the point it is determined by the system range resolution. The vegetation canopy in [93] is modelled as a randomly distributed ensemble of canonical scatterers located above a flat ground. It can be doubted this approach represents the accurate diffuse contribution over real vegetation covers. The present study incorporates realistic canopy structures as described in the previous section to investigate effect of the crop row structure as well as the plant architecture on calculating the diffuse contribution.

Figure 4.9 illustrates the specular and diffuse components of the GNSS-R received power over entire footprint (by using ten Fresnel zones) for the growth stage R1-R4 as a function of zenith observation angle (θ_s). The left most panel shows cross-polarized (RL-RHCP transmit, LHCP receive) received power for 25 dB polarization crosstalk level only (since the effect of crosstalk was minimal for RL-polarization) while the other panels show co-polarized (RR) received power for various polarization crosstalk levels (15 dB, 25 dB,

and 40 dB from left to right) as we will also investigate the effect of antenna impurities on the RR-polarization results. Specular reflection from bare soil is also included as a reference. As a side-note, these received powers for both polarizations seem to have intuitive levels when compared with the previous experimental observations. Because such studies generally analyze reflectivity and/or NBRCS instead of the received power, the measurement ranges for those parameters can be substituted into (4.1) and/or (4.2) in conjunction with the other system parameter values used in these simulations. For instance, Egido et al. [39] performed tower experiments (at 25 m) over a sunflower field and the long-term measured cross-polarized reflectivity was in the range [-20,-5] dB. Such a range for the reflectivity would indicate a range of around [-175,-160] dB for the cross-polarized, coherent received power in this study.

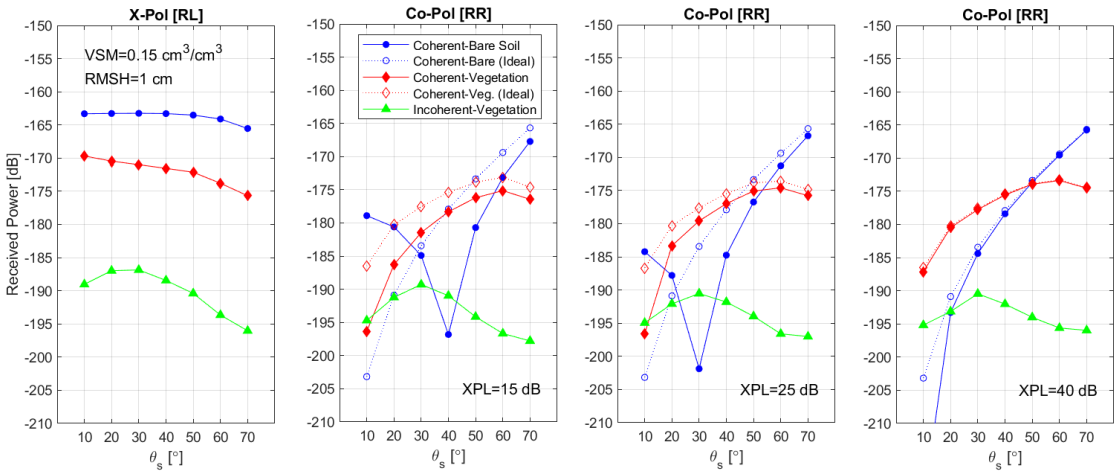

Figure 4.9: Received power coherent and incoherent contributions as a function of zenith observation angle (Equal to the transmitter's incidence angle so that incident wave impinging from the specular direction is considered) for the growth stage R1-R4. EIRP = 27 dB. Receiver Gain = 0 dB (The received power will be modified with actual system EIRP, receiver gain, and processing gains (coherent/incoherent integrations)) (RR: RHCP-RHCP, RL: RHCP-LHCP.

As seen from the left most panel, it is evident that RL-polarized coherent term is much larger than RL-polarized diffuse contribution. Although the results are limited to 0.15 cm^3/cm^3 of VSM, and 1.0 cm of RMSH in the plots, other combinations of the dynamic simulation parameters (growth stages, VSM, RMSH, and θ_s) that are not depicted here show similar results. It can be concluded that the cross-polarized coherent received power always dominates the incoherent one by far in the case that the surface is flat and smooth (RMSH is under 2.0 cm). This observation supports our implications for the homogeneous vegetation case in [93], and it indicates that the incoherent contribution from vegetation over flat and smooth surfaces can be neglected for studies that take only the cross-polarization into account.

The co-polarized received power, on the other hand, shows quite dissimilar behavior compared to the cross-polarized contributions since ground reflection of a circularly polarized wave gets significantly modified in terms of both polarization and strength. For coherent contribution, the co-polarized power is much lower than the cross-polarized one, but both merge as the angle increases and approaches the Brewster angle. The co-polarized incoherent term is mostly lower than all the contributions except lower incidence angles, where it gets comparable to co-polarized coherent contribution. It is also apparent from the figure that the incoherent term experiences smoother changes with angles than the coherent term, which has a drastic dependence on the angle. This could be explained by the fact that volume scattering within vegetation combines two polarized components and reduces their relative strength and angular dependence in diffuse scattering. In addition, ideal (no polarization crosstalk and mismatch) antenna results for co-polarized specular

term for bare soil and vegetated terrains are also included because, due to the simulated actual antenna, results for bare-soil look as if there is a signature at lower angles (the dips in the co-polarized coherent bare soil reflectivity around 40° and 30° for 15 dB and 25 dB crosstalk, respectively). This behavior happens due most likely to the crosstalk between the receiving antenna ports which makes a fraction of the signal go into the opposite polarization port. This could be understood by the interference between the leaked cross-polarized and reduced co-polarized contributions as the signals are summed as complex fields. As also evident, the dip and difference between ideal and simulated actual antenna results disappear with 40 dB crosstalk isolation antenna (the right most result). This indicates that the antenna purity could have a significant effect on co-polarized reflectivity at low incidence angles and a caution must be exercised if co-polarized contributions are utilized in any retrieval algorithms. Hence, the effects of antenna impurities on co-polarized signals are worth being analyzed separately in a future work. There is one more interesting point on the co-polarized received power that in the ideal antenna case, the coherent vegetation term is larger than the coherent bare-soil term at lower angles. The mechanism behind this phenomenon will be explained in detail in the Figure 4.12 discussion, which shows this case more apparently with the help of reflectivity analyses.

Although this study indicates that the incoherent contribution is negligible, a remarkable increase in surface roughness or topography, or change in the vegetation cover or bistatic configuration (such as the receiver altitude compared to vegetation height, or the operating frequency of the antennas) can make it comparable to the coherent contribution for both polarizations. For instance, increasing surface roughness can decrease the coher-

ent contribution and increase the incoherent contribution for both polarizations. Another example can be that a different vegetation cover may lead to a reasonable increase in the diffuse scattering while specular contribution stays somewhat constant. This implies that the co-polarized incoherent contribution can deserve to be accounted for depending on the scene characteristics.

Figure 4.10 and Figure 4.11 are aimed at providing a detailed look at the incoherent contribution alone for varying crop row structures. Figure 4.10 illustrates the effect of relative orientation of corn crop rows (due to changing antenna azimuth observation angles ϕ_s) on both the co-polarized and cross-polarized effective NBRCS ($\underline{\sigma}_e^0$) over the same configuration as in Figure 4.9. The effective NBRCS values are plotted for five different azimuth angles in the range 0°, 30°, 45°, 60°, 90°. It is evident that the azimuth orientation of crop rows makes a slight variation (around 0.2 dB) on the effective NBRCS in both polarizations. Hence, this figure indicates that the orientation of the corn field row structure does not have a critical impact on the incoherent component, when the vegetation architecture (plant body structure) remains the same. This can be attributed to the fact that the diffuse term is driven by the sum of scatterings from individual scatterers within the medium, and large number of scatterers eliminate the azimuthal orientation effect. This behavior resembles measurement of the same average physical property sensed by microwave radiometry as it is not affected by the azimuthal orientation, too. This can be understood that conservation of energy requires a link between emissivity and reflectivity, demonstrating on basic principles that reflectometry and radiometry measurements are related with each other [5].

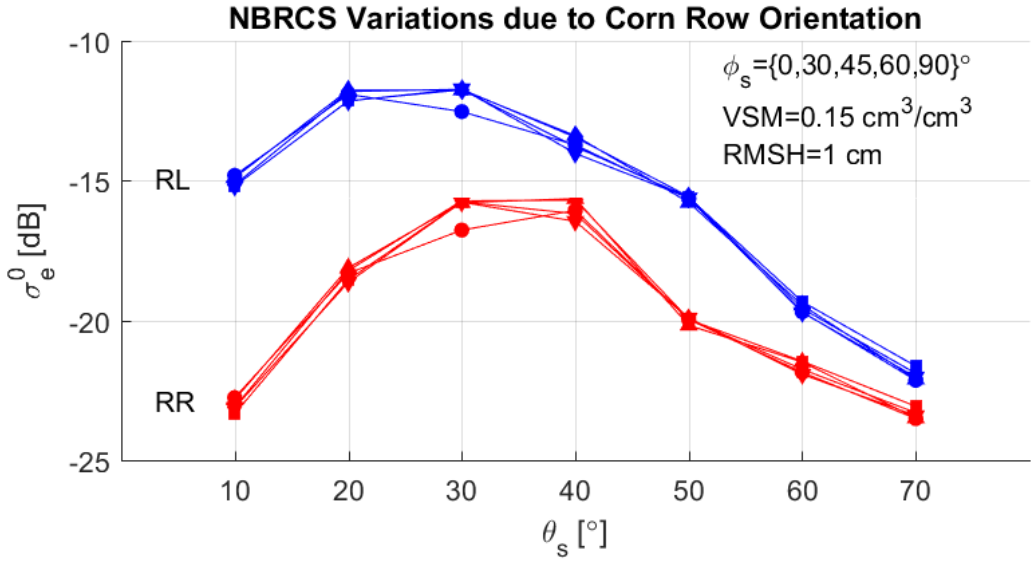

Figure 4.10: Effect of the crop row orientation angles on the effective NBRCS (over the first Fresnel zone) for quasi-periodic seed distribution in the growth stage R1-R4. Each plot in both RR (red) and RL (blue) polarization corresponds to five azimuth angles between antenna observation and the crop rows.

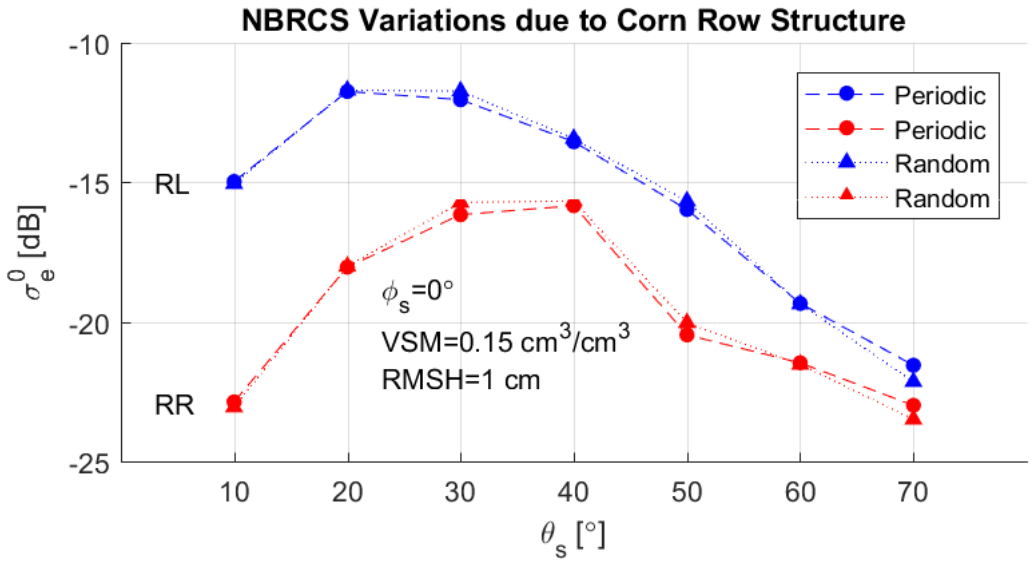

Figure 4.11: Effect of periodic crop row and random seed distribution on the effective NBRCS (over the first Fresnel zone) for both RR and RL polarizations in the growth stage R1-R4.

Figure 4.11 demonstrates a further analysis on the row orientation. Although the main vegetation modeling is based on quasi-periodic seed distribution in our simulations, we also examine a pure periodic row structure with no plant fluctuations in contrast with a random azimuthal distribution of plants, while keeping the average vegetation density constant. In the latter case, the row structure of the corn field is completely discarded; however, individual corn plants still keep their architectural features such as constituent dimensions and orientations. In this case, only an azimuth observation angle of 0° is simulated since the effect of the row orientation angle is found to be negligible in the analysis for Figure 4.10. Figure 4.11 shows that neither periodic, nor random plant distribution can lead to a reasonable difference in the effective NBRCS. Conjunct outcomes of Figure 4.10 and Figure 4.11 imply that diffuse scattering characteristics of the vegetation layer is almost independent from the row structure (orientation and periodicity) while the vegetation architecture remains the same. These conclusions, combined with the observations in Figure 4.9, demonstrate dominance of the cross-polarized coherent term; hence, only the coherent term will be analyzed in the rest of the paper.

Figure 4.12 represents the reflectivity for both bare soil and vegetated fields, VWC, and Leaf Area Index (LAI) as a function of growth stages. The first panel of the figure shows the cross-polarized reflectivity, LAI, and VWC (the right vertical axis). The cross-polarized results are given only at 40° since the effect of angular variation and antenna cross talk were found to be insignificant. On the other hand, the other panels show the co-polarized reflectivity for ideal and simulated actual antenna at various look angles. The purpose of these plots is to describe the relationship between vegetation characteristics

and specular reflectivity. In order to determine the effect of the dielectric constant (which varies much for the leaves and cobs, but not much for the stalks, as shown in Figure 4.8) on the reflectivity, an additional vegetated reflectivity for cross polarization is provided. To perform this, the dielectric constants of the constituents are set to a fixed value $(40 + i10)$, which represents an average value for the entire season. There is no significant difference between the variable and constant permittivity for either cross-polarized or co-polarized reflectivity (the constant dielectric results for co-polarization are not included to keep the plots readable). This can be attributed to the stalk's moisture contents since the stalks are the main contributor to the overall reflectivity, their dielectric constant stays relatively constant, and they retain moisture even in the senescence stage as elaborated in section II-B.

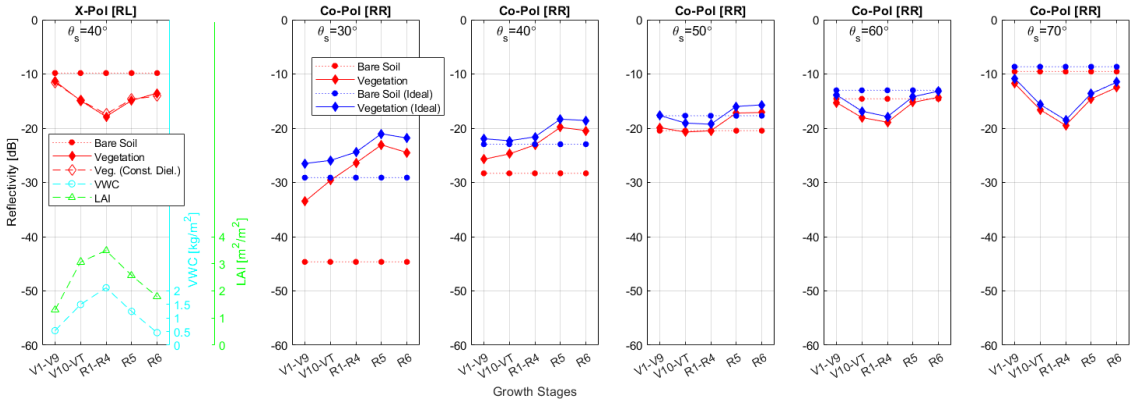

Figure 4.12: Reflectivity, VWC, and LAI as a function of growth stages (VSM $= 0.1cm^3/cm^3$, RMSH $= 1.5cm$). X-pol reflectivity is given only for simulated actual antenna at θ_s = 40° since there is no significant difference due to angle variations and antenna impurities for X-pol. Co-pol reflectivity is given for ideal and simulated actual antenna for θ_s = 30° : 10° : 70° since angle variations and antenna impurities matter for Co-pol. Bare soil reflectivity for both X-pol and Co-pol are given for reference. X-pol reflectivity values for a constant dielectric permittivity $(40 + i10)$ for the entire season are also shown.

As seen in the left most panel of Figure 4.12, the cross-polarized reflectivity seems to be significantly related to the VWC/LAI throughout the growing season. The reflectivity increases as the VWC/LAI decreases, and vice versa. For instance, R1-R4, which is the growth period with the highest VWC/LAI level, has the minimum cross-polarized reflectivity of the season. Similarly, V1-V9 and R6, and V10-VT and R5 exhibits very similar reflectivity trends because they have almost the same VWC/LAI level. This phenomenon can be understood since dielectric permittivity depends on the average water content per volume, but the attenuation is driven by the overall VWC. It is worth repeatedly underlining here that the VWC measurements are not directly involved in the modeling of different growth stages in this study. Instead, the dielectric constant is used to account for separating the dielectric permittivity of the stages. However, the figure shows such a strong correlation between the cross-polarized reflectivity and the *in situ* VWC/LAI measurements for each growth stage. These observations imply that at least the cross-polarized reflectivity is mainly determined by overall VWC/LAI level for a configuration of fixed VSM, RMSH, and zenith observation angle. The dynamic range due to growth stages seems to be around 6 dB in this figure. On the other hand, it should be noted that real-world measurements might not be as sensitive as in these simulated results because of several degrading factors (such as measurement errors and noise) despite exhibiting the same trends. For instance, the tower experiments measured at 25 m [39] showed a VWC sensitivity of 0.3 dB/(Kg/m^2) for a sunflower field.

One can further relate VWC levels per growth stages to Figure 4.5, Figure 4.7, and Figure 4.8 that represent the average constituent dimensions, normalized moisture content

by volume, and average dielectric constants for growth stages, respectively. For instance, VWC levels for V1-V9 and R6 are almost the same because of the fact that plants in the R6 carry as much VWC as V1-V9 thanks to their stalk heights and reasonable amount of average moisture content. In addition, stalks rather than leaves or cobs likely dominate the reflectivity as they have the highest VWC level regarding the three figures again. A further interpretation of the findings of Figure 4.7 and Figure 4.12 together can be that the dielectric constant can be kept constant for investigation involving global land observations, where detailed dielectric information is often lacking.

The cross-polarized reflectivity levels for both bare-soil (about -10 dB) and vegetation cover (roughly from -18 dB to -12 dB) appear to be realistic when compared with the field data measured by Egido et al. [40] from an airborne platform over similar scene conditions. Although it is not directly comparable to the in-season changes in this study, the dynamic range due to varying crop types (such as ploughed, sunflower, and barley) in [40] is at similar levels. On the other hand, Katzberg et al. [82] observed no reflectivity change between the early and the late stages of the corn development. Possible causes of such an experimental observation can be that (1) early and late stages may resemble one of our growth stage pairs that lead to similar reflectivity levels as mentioned above (such as V1-V9 and R6, or V10-VT and R5), (2) noise might have suppressed the reflectivity changes.

The co-polarized reflectivity does not seem to be inversely correlated with VWC and LAI at the low incidence angles as seen in the other five panels in Figure 4.12, where co-polarized reflectivity is plotted for incidence angles at 30° to 70° from left to right. In these

plots, red curves represent the results with the simulated actual antenna (25 dB cross-talk) while the blue curves denote ideal antenna results. At low angles, the discrepancy between ideal and simulated actual antenna is also significant, but as the angle increases, the co-polarization response starts to resemble cross-polarization response and to get correlated with vegetation characteristics. In addition, the effect of antenna cross-talk on the results gets disappeared with higher incidence angles as both co- and cross-polarization reflectivities get within a comparable range of values. Egido et al. [40] showed a quite similar evidence such that the co-polarized reflectivity over a sunflower campaign approaches to the cross-polarized reflectivity as the incidence angle increases. Their co-polarized reflectivity measurements over several bare-soil and crop field campaigns also show that the co-polarized reflectivity values in Figure 4.12 are at similar levels. Another interesting note is that vegetation reflectivity is higher than bare soil reflectivity at lower angles and as the angle increases it gradually goes below the soil reflectivity. For instance, as can be seen at 40°, both cross- and co-polarization vegetation reflectivities (the first and third panels, respectively) are sandwiched between bare soil reflectivities. This may seem counter intuitive at the first look. However, this behavior can be explained with the fact that at lower angles, the ground reflection is more in favor of reversing RHCP to LHCP, but LHCP and RHCP merge as the angle increases and vegetation tends to disturb polarization difference of incoming and reflected signals, particularly at lower angles. In other words, vegetation tends to smooth out the polarization difference as this is more apparent at lower incidence angles when the ground reflectivities are very distinct at both polarizations. This means that the vegetation layer does not only make an attenuation effect on the signal power, but

it also changes the phase of the signals. For cross-polarized term, there is no surprise that both the attenuation and phase change due to vegetation layer will make the vegetated term lower than the bare-soil. However, for co-polarized case, at lower angles, the bare-soil term has the lowest reflectivity levels because the ground reflection reverses the RHCP to LHCP. As some vegetation layer is added onto the ground, it will disturb both the RHCP polarization of incoming signals and LHCP polarization of reflected signals and will eventually lead to higher reflectivity values than the bare-soil.

Figure 4.13 and Figure 4.14 are dedicated to disclosing the impact of VSM and RMSH on the reflectivity. Figure 4.13 demonstrates the cross- and co-polarized reflectivity as a function of VSM in conjunction with the growth stages, when the incidence angle is 30° and RMSH is 1.0 cm. While the left most panel shows RL-polarized reflectivity, the other three panels show the RR-polarized reflectivity for different antenna cross-talk polarization isolation levels (15 dB, 25 dB, and 40 dB from left to right). As seen from Figure 4.13, RL-polarized reflectivity increases with increasing VSM levels and seems to reach a saturation when the VSM level is sufficiently large. Therefore, VSM increase of 0.5 cm^3/cm^3 can only raise the reflectivity by about 0.1 dB for higher VSM levels, whereas the same change can lead to a jump of around 2 dB for lower levels (Total variation of reflectivity from dry to wet SM is about 6 dB, which is significant). A major implication of this can be a reduced GNSS-R sensitivity to high VSM levels (such as above 0.40 cm^3/cm^3) for cross-polarization. On the other hand, one interferometric experiment [97] found distinguishing between lower VSM values (such as 0.05 and 0.10 cm^3/cm^3) to be hard. They also make an important point that the decrease in penetration depth reduces much after heavy rainfall.

In such a scenario, the observed values might be lower than the actual VSM levels. This indicates that an inversion model that is capable of finding and handling rapid changes in VSM levels should be developed to overcome such a problem.

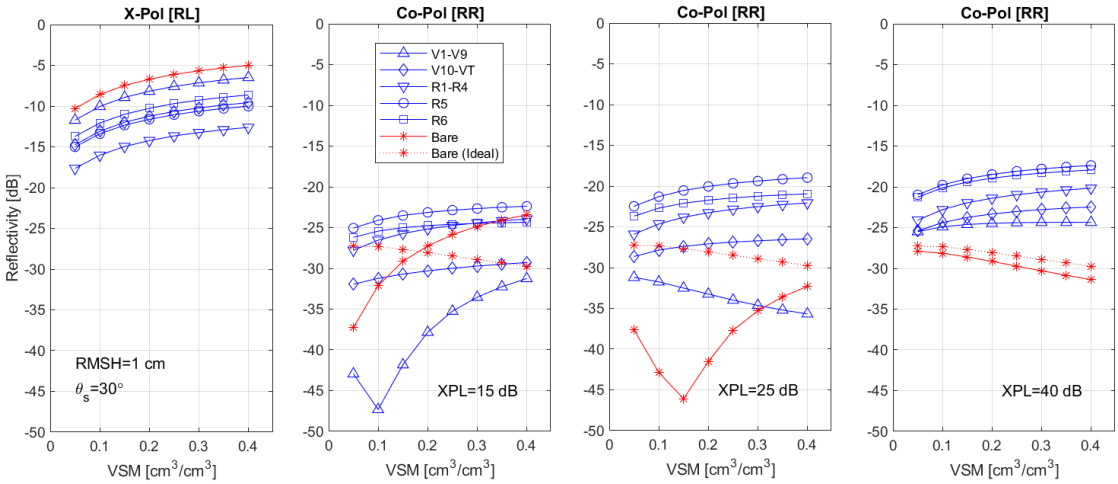

Figure 4.13: Reflectivity as a function of VSM and growth stages. Bare soil reflectivity for both the ideal and simulated actual antenna cases given for reference.

RL-polarized reflectivity characteristics of the distinct growth stages are apparent as in Figure 4.12. For instance, R5 and V10-VT has almost the same reflectivity for all VSM levels. In general, growth stages do not make an impact on SM sensitivity but instead introduces an offset. Equal sensitivity to VSM variations for different levels of vegetation development (growth stages) is due the fact that the specular vegetation reflectivity involves wave trasmissivity which multiplies the ground reflectivity (i.e., adds trasmissivity in the logarithmic scale) as well as the fact that no measurement noise is considered in this study. Therefore, vegetation attenuation only introduces an offset. Under real conditions where noise exists, the estimation of ground reflectivity under dense vegetation will

get harder and ground sensitivity will be likely compromised with the other errors such as vegetation attenuation estimation. The effect of VSM on the co-polarized reflectivity generally looks similar to the cross-polarized case for high vegetation, but the effect of antenna impurities such as cross-talk polarization leakage is significant on the light vegetation and bare soil conditions, where there are some abrupt behavior between 0.1 to 0.2 VSM for 15 dB and 25 dB isolation. Similar to the co-polarized results presented in Figure 4.9, abrupt behavior can be attributed that the interference between the leaked cross-pol and reduced co-polarized contributions as the signals are summed as complex fields. This behavior disappears at 40 dB cross-polarization isolation and the results converge to the ideal antenna results. Egido et al. [40] measured the ratio of cross-polarized reflectivity to co-polarized reflectivity as a function of VSM over bare-soil fields. Their observation can be related to our cross-polarized (the left-most panel) and co-polarized with 40 dB-XPL (the right-most panel) simulations since they selected the receiving antennas' XPL to exceed 25 dB at boresight to prevent polarization leakage. They observed a linear increase in the polarization ratio, which is also the case for our bare-soil simulations. Tower experiments over a sunflower field [39] showed a similar fashion between the cross-polarized and co-polarized reflectivity measurements. Potential future work will attempt to increase the understanding of the abrupt co-polarized behaviors that are attributed to the antenna impurities in this paper.

Figure 4.14 additionally introduces the effect of RMSH on the reflectivity for a zenith observation angle of 40° and the R5 stage. For both co-polarization and cross-polarization, RMSH has a minimal effect on the reflectivty. For instance, change in RMSH from 0.5cm

to 1.5 cm produces 1-2 dB variations in the reflectivity as seen from the figure. Moreover, there is no sign of a saturation for the impact of RMSH. However, it should be kept in mind that the SCoBi model processes all these RMSH values as smooth roughness values for a flat surface [93]. Therefore, sufficiently high roughness levels and high topographic relief could dramatically change the results.

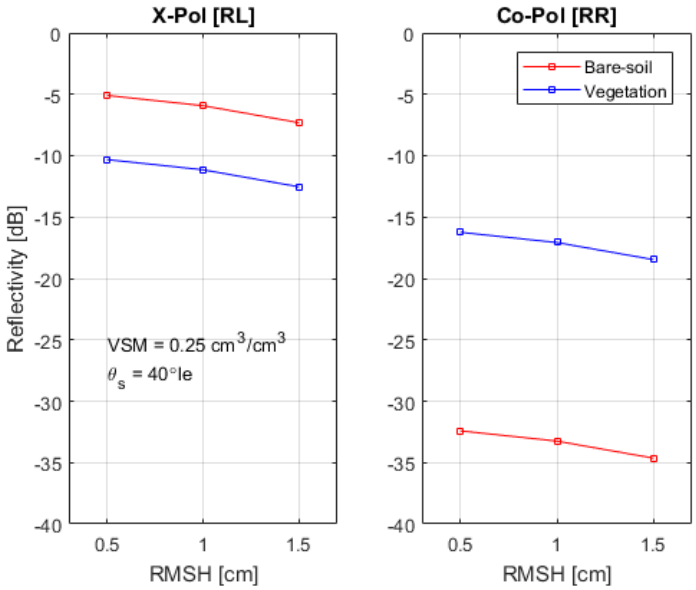

Figure 4.14: Reflectivity as a function of VSM and RMSH variations for the growth stage R5.

4.4 Conclusion

In this paper, we presented the quantitative findings for the reflectometric effects of land geophysical parameters such as soil moisture, surface roughness, and vegetation water content by using our recently developed scattering model, SCoBi-Veg. We detected a set

of significant GNSS-R signatures over the growing season of a row crop corn field that is simulated by using *in situ* measurement data as follows:

- We demonstrated the dominance of the coherent contribution in both co-polarized and cross-polarized received powers for a flat and smooth surface case.
- We examined the impacts of possible row orientation variations and plant distributions on the effective NBRCS in order to check if the diffuse contribution would dominate the specular contribution in any case. Results show that the changes in the row structure without any modification in the plant architectures have minimal consequences on the effective NBRCS. These initial findings indicate that the coherent contribution can be considered alone for land applications if, the ground can be assumed as flat and smooth, and if the vegetation layer height is much lower than the receiver altitude.
- We analyzed the reflectivity and VWC as a function of growth stages, and we found strong correlation between these two for cross-polarization. Furthermore, using a constant dielectric permittivity for each plant constituent through the entire season does not lead to a significant change on the reflectivity. This shows that the changing VWC levels highly dominate the cross-polarized reflectivity; however, changing dielectric constants have a limited effect. This is because the dielectric permittivity is determined by the average volumetric water content, whereas the attenuation is driven by the overall VWC. This can help the researchers assume a reasonably fixed dielectric constant value to model vegetation covers.
- We presented the quantitative effects of VSM and RMSH variations on the reflectivity. The total variation of reflectivity from dry to wet SM is about 6 dB and VSM sensitivity staturates at higher VSM values while RMSH has a minimal impact on reflectivity for moderately rough surfaces.
- Finally, it is demonstrated that the antenna cross-talk leakage could have a significant effect on co-polarized reflectivity at low incidence angles and a caution must be exercised if co-polarized contributions are utilized in any retrieval algorithms.

The simulated studies via forward models like those presented in this paper can help develop inversion models for estimating the land geophysical parameters, if they can be expanded for various configurations. As a result, modeling and quantification studies as performed in this study have a potential to clarify experimental observations and possible ambiguities.

CHAPTER 5

HIGH SPATIO-TEMPORAL RESOLUTION CYGNSS SOIL MOISTURE ESTIMATES USING ARTIFICIAL NEURAL NETWORKS

This chapter is inherited from the following publication:

O. Eroglu, M. Kurum, D. Boyd, and Ali C. Gurbuz, "High spatio-temporal resolution CYGNSS soil moisture estimates using artificial neural networks," *Remote Sensing*, vol. 11, no. 19, 2019, pp. 1-32.

Significant author contributions to the above publication were as follows: Conceptualization and Methodology by **Eroglu**, Kurum, and Gurbuz; Investigation and Validation by **Eroglu**, Kurum, and Gurbuz; Writing by **Eroglu**; Review and Editing by **Eroglu**, Kurum, Boyd, and Gurbuz; Visualization by **Eroglu**; Supervision by Kurum; Project Administration by **Eroglu** and Kurum; Funding acquisition by **Eroglu** and Kurum.

This paper presents a learning-based, physics-aware soil moisture (SM) retrieval algorithm for NASA's Cyclone Global Navigation Satellite System (CYGNSS) mission. The goal of the proposed novel method is to advance CYGNSS-based SM estimations, exploiting the spatio-temporal resolution of the GNSS reflectometry (GNSS-R) signals to its highest potential within a machine learning framework. The methodology employs a fully connected Artificial Neural Network (ANN) regression model to perform SM predictions

through learning the non-linear relations of SM and other land geophysical parameters to the CYGNSS observables. *In-situ* SM measurements from several International SM Network (ISMN) sites are used as reference labels; CYGNSS incidence angles, derived reflectivity and trailing edge slope (TES) values, as well as ancillary data are exploited as input features for training and validation of the ANN model. In particular, the utilized ancillary data consist of normalized difference vegetation index (NDVI), vegetation water content (VWC), terrain elevation, terrain slope, and h-parameter (surface roughness). Land cover classification and inland water body masks are also used for the intermediate derivations and quality control purposes. The proposed algorithm assumes uniform SM over a $0.0833^{\circ} \times 0.0833^{\circ}$ (approximately 9 km $\times$ 9 km around the equator) lat/lon grid for any CYGNSS observation that falls within this window. The proposed technique is capable of generating sub-daily and high-resolution SM predictions as it does not rely on time-series or spatial averaging of the CYGNSS observations. Once trained on the data from ISMN sites, the model is independent of other SM sources for retrieval. The estimation results obtained over unseen test data are promising: SM predictions with an unbiased root mean squared error of 0.0544 cm^3/cm^3 and Pearson correlation coefficient of 0.9009 are reported for years 2017 and 2018.

5.1 Introduction

Soil moisture (SM) has an active role in the Earth's water cycle between the ground and the air. This role makes SM a key land geophysical parameter for understanding hydrologic processes, vegetation states, and climatic conditions in order to improve applications such

as hydrologic modeling, agriculture, crop yield estimation and vegetation change detection, as well as weather and climate forecasts [44, 41, 103]. Global SM retrieval at high spatio-temporal resolutions, therefore, has been an important research topic for the past several decades.

The current state of the science for global SM estimation relies on microwave remote sensing with the use of traditional instruments such as monostatic radars and radiometers. This is because the microwave frequencies are sensitive to the changes in the soil dielectric properties with respect to the presence of moisture content [35]. In particular, L-band radiometry is commonly used because it has an increased sensitivity to the near-surface SM (0-5 cm) as well as its reduced attenuation due to the atmospheric losses, surface roughness, and vegetation cover [14]. ESA's Soil Moisture and Ocean Salinity (SMOS) [86] and NASA's Soil Moisture Active Passive (SMAP) [42] satellites are current missions that have on-board L-band radiometers and provide SM retrievals with a spatial resolution of around 40 km and revisit time of 2-3 days. On the other hand, radar backscattering generally offers finer spatial resolutions (tens of meters to few km) within longer revisit times except SMAP's radar instrument (L-band), which was capable of providing providing a spatial resolution of 3 km and a revisit time of 2-3 days with the help of its rotating antenna before a hardware failure in mid 2015. SMAP mission was designed to make use of a 6 m mesh reflector antenna for both radar and radiometer instruments to provide high spatio-temporal resolution SM products [42, 88]. In addition, the radar backscattering data of ESA's Sentinel-1 (C-band) [126] and DLR's TERRASAR-X (X-band) [6] were used for global SM estimates.

Global Navigation Satellite System Reflectometry (GNSS-R) is an alternative microwave remote sensing approach, which is based on reception of the reflected GNSS signals from the Earth surface in a bistatic geometry [178]. This approach shows a great potential for remote sensing of SM because it operates at L-band. Moreover, it can offer high spatial resolutions with low revisit times by using constellations of small satellites due to being strictly receive-only [142]. GNSS-R applications have seen advancements for various Earth science areas over the past two decades that have resulted in the launch of new satellite missions [109, 18, 16, 21]. For instance, the first dedicated spaceborne GNSS-R receiver was a secondary payload on-board the UK Disaster Monitoring Constellation (DMC) [162]. It has demonstrated the potentiality of GNSS-R for the remote sensing of ocean, ice, and land geophysical parameters [60]. The UK Technology Demonstration Satellite (TDS-1) was launched in 2014 with an improved primary GNSS-R payload which provided more data that were used to study GNSS-R sensitivity to SM [13, 23]. NASA's Cyclone GNSS (CYGNSS) was launched in December 2016 to improve weather predictions by estimating ocean winds between 38° north and 38° south latitudes [141]. CYGNSS has eight small satellites in orbit, each with four channels, allowing simultaneous measurements from up to 32 channels. It has a mean revisit time of seven hours over the ocean. The key orbital and instrumental specifications of the CYGNSS mission are listed in Table 5.1. The constellation records a considerable amount of land observation data as well. CYGNSS measurement sensitivity to the surface SM has been reported by multiple studies [142, 24, 19, 2]. These efforts demonstrated GNSS-R's potential to com-

plement the traditional passive and active instruments for monitoring surface SM at global scales for improved spatio-temporal resolutions.

Table 5.1: Orbital and instrumental specifications of CYGNSS [138, 24]

Parameter	**Value**
Orbit	Non-synchronous
Altitude	$\sim$ 520 km
Period	95 minutes
Spatial Resolution	$\sim 25\,km \times 25\,km$ (incoherent), $\sim 0.6\,km \times 6.6\,km$ (coherent)
Revisit Times	3 hours median, 7 hours mean (over the ocean)
Frequency	1.57542 GHz
Polarization	LHCP (Down-looking GNSS-R antenna)

CYGNSS is not designed for land observations; however, a CYGNSS-based, accurate SM retrieval algorithm could enable scientists **(i)** to specify the requirements for dedicated SM missions of the future, **(ii)** to create new algorithms utilizing existing land data, and **(iii)** to discover new calibration/validation approaches for dedicated GNSS-R SM missions. Motivated by this, there have been increasing efforts to develop SM retrieval algorithms for CYGNSS observations. For example, Chew and Small [24] correlated the changes in the CYGNSS signal-to-noise ratio (SNR) with the SMAP SM estimations, assuming that CYGNSS land measurements are dominated by the coherent reflections. They used mean SMAP SM values as reference with these correlations to obtain daily, CYGNSS-based SM estimations from SNR changes for each SMAP Equal-Area Scalable Earth (EASE) grid (36 km $\times$ 36 km) [11]. The overall unbiased root-mean-squared error (ubRMSE) of their algorithm is 0.0450 cm^3/cm^3. Although the estimation method itself is relied on SMAP

SM data as a reference, its significant benefit was that linking CYGNSS SNR to SMAP SM products allowed the use of CYGNSS observations to fill in the gaps between the adjacent SMAP observations. Kim and Lakshmi [87] introduced a relative SNR (rSNR) and SM derivation from CYGNSS delay-Doppler maps (DDM) to infill the gap between adjacent SMAP revisits. They reprojected the CYGNSS SNR observations into SMAP's 9-km EASE grids and calculated the average of these grids to acquire daily SM estimations. They combined rSNR with SMAP SM values to acquire daily SM estimations over high vegetation density as well. They reported correlation results at useful levels (Pearson R of 0.77 between CYGNSS-derived SM and SMAP) over moderate vegetation density but with reduced correlations (R=0.68) over dense vegetation; however, they did not report any error level (such as ubRMSE). Carreno-Luengo et al. [19] did not propose a CYGNSS-based SM estimation method; however, they made use of an approximated CYGNSS reflectivity, which is the ratio of the calibrated reflected and direct SNR measurements, assuming predominant coherent reflections over land. They linked CYGNSS reflectivity approximation to the SM changes for several land cover types. Clarizia et al. [28] introduced the trilinear regression-based reflectivity-vegetation-roughness (R-V-R) algorithm that derives daily SM estimations at a 36 km $\times$ 36 km resolution as a function of the CYGNSS reflectivity as well as SMAP vegetation opacity and roughness coefficient. The algorithm was developed considering the dominance of the coherent reflections over land. The R-V-R performance was compared globally to the SMAP SM product and reported to have a RMSE of 0.07 cm^3/cm^3. Al-Khaldi et al. [2] proposed a time-series SM retrieval algorithm that produces 3-day and 1-day SM estimates. In contrast to former studies, they assumed that the

CYGNSS land returns are mostly driven by the incoherent scattering unless inland water bodies exist within the footprint. Therefore, they used the CYGNSS normalized bistatic radar cross section (NBRCS) and mean-square slope (MSS) instead of the DDM SNR. As a result, they provided SM estimations at relatively coarse resolutions ($0.2^\circ \times 0.2^\circ$ lat/lon grid roughly 22 km around the equator) with an overall RMSE of 0.04 cm^3/cm^3. They constructed a system of equations for 30-day time-series of the CYGNSS measurements, which is indeed an under-determined system with 29 equations. Hence, they incorporated the SMAP maximum and minimum SM values into the algorithm for bounding the system. They also assumed that changes in vegetation and surface roughness occur much slower compared to changes in SM.

This paper proposes a physics-aware machine learning approach through capturing the non-linear dependencies of the CYGNSS observables to SM values and several bio/geophysical parameters. An Artificial Neural Network (ANN) is employed to learn the complex non-linear relations. The term ”physics-aware” in this manuscript refers to the use of several ancillary data sets and International Soil Moisture Network (ISMN) measurements to represent the vegetation and ground dynamics in the learning process. The details of the data usage will be explained in the next section. Daily SM measurements are used both in training of the model and validation of the SM predictions. One of the main objectives of this study is to initiate a novel, practically applicable SM retrieval algorithm that can provide sub-daily SM products within few kilometers by utilizing the individual CYGNSS observations as the algorithm inputs. Once trained on a reference data set, this method neither requires SM information from other satellite missions, nor operates on spatial or temporal

averaging of the CYGNSS observations. SM retrieval performances for multiple ISMN sites are visually and quantitatively demonstrated.

The rest of the paper is organized as follows: Section 5.2 describes the theoretical background of the GNSS-R based SM retrieval, possible use of CYGNSS data products, and the related challenges, followed by Section 5.3 with the explanation of the SM retrieval methodology, details of the ANN model, acquisition of the data sets, as well as training and validation. Section 5.4 provides the SM estimation results along with the visualizations and statistical performance metrics achieved. Section 5.5 gives a comprehensive discussion of the findings and points to be improved in the future. Section 5.6 concludes the study.

5.2 Theoretical Background

Bistatic CYGNSS radars receive the L-band GNSS-R signals that are transmitted by the GPS satellites and subsequently forward-scattered from Earth's surface in the specular direction. This configuration of the CYGNSS and GPS constellations functions as a bistatic radar at L-band, which receives information relevant to the scattering surface properties. SM can be retrieved as part of such information overland as it is the primary determinant of the dielectric constant of the scattering surface. This section provides theoretical background for SM retrieval from bistatic radar observations, potential use of CYGNSS observables in such a task, and related challenges.

5.2.1 Inversion of The Bistatic Radar Equations

An ideal GNSS-R based SM retrieval approach would rely on inversion of the bistatic radar equations to acquire the surface reflectivity. The surface reflectivity would be cor-

rected for the vegetation cover and surface roughness effects to obtain Fresnel reflection coefficient. Fresnel reflection coefficient could then be related to SM with the help of Fresnel reflection equations.

For cases where specular reflections are fully dominant, the coherent component of the bistatic received power can be written as follows [33, 178, 167, 93]:

$$P_{RL}^{coh} = \left(\frac{\lambda}{4\pi}\right)^2 \frac{P_t G_t G_r}{(r_{st} + r_{sr})^2} \Gamma_{RL}(\theta_i) \tag{5.1}$$

where P_{RL}^{coh} denotes the coherently received power. The subscripts R and L stand for the right-hand circularly polarized (RHCP) GNSS transmit antenna and the left-hand circularly polarized (LHCP) downward-looking GNSS-R antenna, respectively [178]. λ is the free space wavelength, P_t is the peak power of the transmitted GNSS signals, G_t is the gain of the transmitter antenna, G_r is the gain of the receiver antenna. r_{st} is the distance between the specular reflection point and the GNSS transmitter, while r_{sr} is the distance between the specular reflection point and the GNSS-R receiver. $\Gamma_{RL}(\theta_i)$ denotes the specular reflectivity at a local incidence angle of θ_i.

The incoherent component of the bistatic received power can be written as follows [179, 139]:

$$P_{RL}^{inc} = \frac{(\lambda)^2}{(4\pi)^3} \frac{P_t G_t G_r}{r_{st}^2 r_{sr}^2} \sigma_{RL} \tag{5.2}$$

where P_{RL}^{inc} denotes the bistatic received power due to the the diffuse scattering over the surface. σ_{RL} is the bistatic radar cross section (BRCS) in m^2. BRCS can be further defined as follows:

$$\sigma_{RL} = A_s \sigma^0_{RL} \tag{5.3}$$

where, the quantity A_s is the contributing surface area (frequently called as the glistening zone [178]), where the diffuse scattering originates from. σ^0_{RL} is the normalized BRCS (NBRCS), which includes the spreading loss and the path-dependent phase terms for diffuse mechanisms (such as single scattering or multi-scattering) in the various multi-path directions for the scattering particles (mainly due to the vegetation canopy) and the surfaces (topography and roughness) [93].

The bistatic received signals are assumed to be dominated by the coherent reflections when the surface is relatively flat (no topographic relief) and smooth (weak roughness), having no or non-heavy vegetation cover [142, 178, 24, 28, 51]. The surface reflectivity in this case can be obtained by directly solving (5.1) for $\Gamma_{RL}(\theta_i)$, as shown below:

$$\Gamma_{RL}(\theta_i) = \left(\frac{4\pi}{\lambda}\right)^2 \frac{P^{coh}_{RL}(r_{st} + r_{sr})^2}{P_t G_t G_r} \tag{5.4}$$

Furthermore, it can also be computed by substituting (5.1) into (5.2) with $P^{coh}_{RL} = P^{inc}_{RL}$ (i.e. equating the right-hand-sides of two equations to each other) and obtaining $\Gamma_{RL}(\theta_i)$ as a function of σ_{RL} [135] as follows:

$$\Gamma_{RL}(\theta_i) = \frac{\sigma_{RL}(r_{st} + r_{sr})^2}{(4\pi) r^2_{st} r^2_{sr}} \tag{5.5}$$

Above Γ_{RL} calculation (correction for the path loss and the 4π term) functions as a correction to σ_{RL} for coherency assumption, whereas σ_{RL} is originally computed assuming incoherency.

After obtaining the surface reflectivity, $\Gamma_{RL}(\theta_i)$, by using either (5.4) or (5.5), the Fresnel reflection coefficient, $\Re_{RL}(\theta_i)$, should be derived from $\Gamma_{RL}(\theta_i)$ for SM retrieval. This is because $\Re_{RL}(\theta_i)$ is mainly driven by the moisture content of the soil (SM) [35]. $\Re_{RL}(\theta_i)$ can be calculated by correcting $\Gamma_{RL}(\theta_i)$ for the vegetation [74] and surface roughness effects assuming that the rough surface under the vegetation to be flat and smooth and to follow Kirchhoff's approximation with a Gaussian height distribution [26] as follows:

$$\Gamma_{RL}(\theta_i) = |\Re_{RL}(\theta_i)|^2 \gamma^2 exp(-hcos^2(\theta_i)) \tag{5.6}$$

where, the exponential term in (5.6) accounts for the surface roughness effects. The *h-parameter* is assumed linearly related to the root-mean-square-height surface roughness [124], as follows:

$$h = 2k_0 s \tag{5.7}$$

where k_0 is the angular wavenumber and s is the surface root-mean-squared (rms) height.

The squared transmissivity, γ^2, in (5.6) accounts for the wave attenuation as the waves propagate from the top of the vegetation canopy to the ground and then from the ground to the top of the vegetation cover again. The transmissivity depends on the vegetation optical depth, τ, and the incidence angle as follows:

$$\gamma = exp(-\tau sec(\theta_i)) \tag{5.8}$$

γ and τ (and parameter b in (5.9)) are dependent on the electromagnetic signals' polarization, but the polarization notation is waived here for simplicity. The vegetation optical depth has been previously related to vegetation water content (VWC) and a land cover-based proportionality value (b) that depends on both the vegetation structure and the microwave frequency in the literature [74], and this approach has been successfully applied to the coarse spatial resolution SMOS/SMAP missions [86, 42], as shown below:

$$\tau = b \times VWC \tag{5.9}$$

VWC was empirically derived from normalized difference vegetation index (NDVI) by the SMAP mission with additional utilization of the minimum and maximum NDVI values of ten-year time-series, and the *stem factor* parameter that comes from a land cover-based lookup table (LUT) [22] as follows:

$$VWC = (1.9134 \times NDVI^2 - 0.3215 \times NDVI) + stem factor \times \frac{NDVI_{max} - NDVI_{min}}{1 - NDVI_{min}} \tag{5.10}$$

Equation (5.6) could be solved for the Fresnel reflection coefficient, $\Re_{RL}(\theta_i)$, substituting equations from (5.6) to (5.10). It could then be related to the soil dielectric constant, ϵ_r, with the help of the Fresnel reflection equations as follows:

$$\Re_{RL}(\theta_i) = \frac{1}{2}(\Re_{VV}(\theta_i) - \Re_{HH}(\theta_i)) \tag{5.11}$$

where

$$\Re_{HH}(\theta_i) = \frac{cos\theta_i - \sqrt{\epsilon_r - sin^2\theta_i}}{cos\theta_i + \sqrt{\epsilon_r - sin^2\theta_i}} \tag{5.12}$$

$$\Re_{VV}(\theta_i) = \frac{\epsilon_r cos\theta_i - \sqrt{\epsilon_r - sin^2\theta_i}}{\epsilon_r cos\theta_i + \sqrt{\epsilon_r - sin^2\theta_i}} \tag{5.13}$$

The soil dielectric constant, ϵ_r, can be related to SM with the help of a ground dielectric mixing model by using soil texture information. A number of dielectric mixing models has been developed in the literature such as Dobson [35], Mironov [115], or Wang-Schmugge [170] models. It should be noted that some of these models might require the use of additional geophysical parameters such as soil temperature [115].

5.2.2 Potential Use of CYGNSS Data

CYGNSS receivers process delay-Doppler maps (DDM) as the main observatory product [178]. CYGNSS Level 1 v2.1 Science Data Products, definitions of which can be found in Appendix E, include a number of geometry- and instrument-related DDM-derived variables.

With respect to the consideration of the dominant coherent reflections described previously, the surface reflectivity can be approximated by using either (5.4) or (5.5) with the CYGNSS data products. For instance, the CYGNSS data can be substituted into the calibration parameters in either equation as follows: $gps\_tx\_power\_db\_w$ for P_t, $gps\_ant\_gain\_db\_i$

for G_t, sp_rx_gain for G_r, $tx\_to\_sp\_range$ for r_{st}, and $rx\_to\_sp\_range$ for r_{sr}. In order to perform the calculation by using (5.4), the bistatic received power, P_{RL}^{coh} is needed to be substituted by a CYGNSS observation. Using either ddm_snr or peak of the $power\_analog$ DDM was investigated by previous studies [24, 87, 19, 28, 135]. The ddm_snr product accounts for the peak DDM signal-to-noise ratio (SNR) and is computed as $10log_{10}(S_{max}/N_{avg})$, where S_{max} is the maximum value (in raw counts) in a single DDM bin and N_{avg} is the average raw noise counts per bin [139].

To solve (5.5) for the surface reflectivity, *brcs* data product (Appendix E) of the CYGNSS mission can be used. In principle, this should produce an output equal to the use of peak power in (5.4); however, the resulting reflectivity approximations have differences from each other, which is most likely due to the internal calibration process when generating *brcs*. Although the CYGNSS data products are originally calibrated for ocean surface sensing, using *brcs* over land is valid since it is only calibrated for the instrumental and geometric parameters [135]. *brcs* is published as a DDM within the CYGNSS Level 1 data; however, the peak value can be exploited under the coherency assumption. The rest of the derivation through (5.5) is based on the calibrations with respect to the range terms.

5.2.3 Challenges

Despite the potential usability of the CYGNSS data products in the reflectivity calculations through (5.4) or (5.5), the uncertainties in the determination of these data products in the present Level 1 data version (v2.1) would introduce errors in the estimations. Uncertainties in the current CYGNSS data include estimation of the receiver gain as well

as the GPS transmitted power and gain. Furthermore, the transmitter and the receiver ranges to the specular point (SP) might involve errors since the current SP calculation of the CYGNSS mission uses an ellipsoidal model of the Earth, ignoring topography over land [62, 28]. Since the equations from (5.4) to (5.11) can only provide the optimal solution with the computation of an absolute reflectivity, such errors would require subsequent corrections for accurate SM retrieval.

Additionally, varying land covers (especially, mixtures of heavy vegetation canopies such as forests) and topographic relief over land can introduce an ambiguity about where and under what conditions the coherent reflection regime is dominant. When the incoherent component of the bistatic received power superimposes or dominates the coherent reflections, the use of (5.4) and (5.5) would lead to inaccuracies.

In addition to the aforementioned challenges so far, the SM retrieval process itself contains high complexity and non-linearity. This is because equations (5.4) through (5.11) imply that the retrieval problem is dependent not only on the reflectivity and SM but also on the vegetation, surface roughness, topography, and soil texture through a combination of linear and non-linear relations. Moreover, these land geophysical parameters have distinct variability ranges. In addition, the CYGNSS DDM instrument (DDMI) has diverse sensitivities to these parameters [157]. As a result, CYGNSS observations exhibit non-linear relations with the dynamic land parameters. This leads to parameter ambiguity, where varying combinations of multiple land geophysical parameters might result in the same or close sensor measurements. Parameter ambiguity makes the SM retrieval an ill-posed problem. Additionally, to obtain accurate retrieval results, the impact of the measurement

geometry (incidence angle) as well as the internal and external noise need to be properly accounted for. Previous airborne GNSS-R experiments and modeling studies reported supporting observations and simulated results for these effects, where a dynamic range of roughly 15 dB is determined jointly by the several dynamic geophysical parameters throughout crop seasons [39, 51].

The CYGNSS constellation has sub-daily, quasi-random observations with fine spatial resolutions. Despite the advantage due to high spatio-temporal resolution, this complicates the use of ancillary data for accurate representation of the land geophysical parameters. In other words, finding sufficiently accurate input parameters at the high spatio-temporal resolutions of the CYGNSS observations to correct the vegetation and surface roughness effects as well as to solve the Fresnel reflection equations is of concern. It is evident from equations (5.4) through (5.11) that knowledge of the various parameters from vegetation and surface roughness to soil texture is needed at CYGNSS's resolution for accurate retrieval using the given physical model for the coherent reflection assumption. In fact, simplified LUTs (such as land cover-based or globally constant values) or average values per coarse grids were used to approximate these parameters in the coarse spatial resolution SM retrieval missions such as SMOS [84] and SMAP [124]. The reduced sensitivity of the radiometers to the roughness and vegetation makes this possible for the relatively coarse observations of these missions. Nonetheless, CYGNSS provides quite fine spatial resolutions (from hundreds of meters to several kilometers, depending on the coherence, incidence angle, elevation, and orientation) with frequent revisit times (several hours to few days), and its measurements are highly sensitive to the topography, surface roughness,

and vegetation changes [19]. Thus, even successive observations along the CYGNSS track can have largely different values due to the spatial variations in these land geophysical parameters. Figure 5.1 illustrates this phenomenon by inspecting the CYGNSS observations after the Hurricane Florence landfall on North Carolina, USA. It shows the mean of the uncalibrated CYGNSS SNR values from 14-18 September 2018 that are averaged per SMAP grid pixels (Roughly $36\,km \times 36\,km$). The zoomed-in version of one of the grids demonstrates the actual CYGNSS data where spatial and temporal variability of the CYGNSS measurements even within a SMAP pixel is apparent. It is evident from Figure 5.1 that the CYGNSS mission, or GNSS-R in general, offers a sufficiently high spatio-temporal resolution which can help improve hydrological and agricultural applications. Therefore, the detailed information from this resolution gets lost due to any spatial gridding and/or temporal averaging while developing a CYGNSS-based SM retrieval methodology. For instance, previous SM retrieval attempts gridded multiple CYGNSS data points into larger grids (such as $36\,km \times 36\,km$ SMAP EASE-grid) even though the coherent reflections over land are considered [24, 28].

Regarding the methodological challenges and the retrieval complexities described so far, regression techniques can be practical for the CYGNSS-based SM retrieval problem instead of pure explicit solution of the physical model shown in Section 5.2. In principle, such techniques are based on fitting a regression model between the known SM values from a reference data set (such as SMAP, SMOS, or *in situ* SM networks) and the CYGNSS observations (possibly in conjunction with ancillary data), and exploiting this model to perform future SM estimations. There have been previous efforts conducted to

obtain variations of linear regression models [24, 28]. As Clarizia et al. [28] state, however, linear regression approaches may be too simplistic to deal with the non-linear dependence of the CYGNSS observations on SM and the other land geophysical parameters (SM, vegetation canopy, topography, surface roughness, and soil texture). For instance, large local variations between NDVI and topography occur at very high resolutions (few tens of meters) [176]. Such a high spatial variation of parameter correlations, combined with diverse sensitivity of CYGNSS DDMI to different parameters, would make linear regression approaches perform poorly.

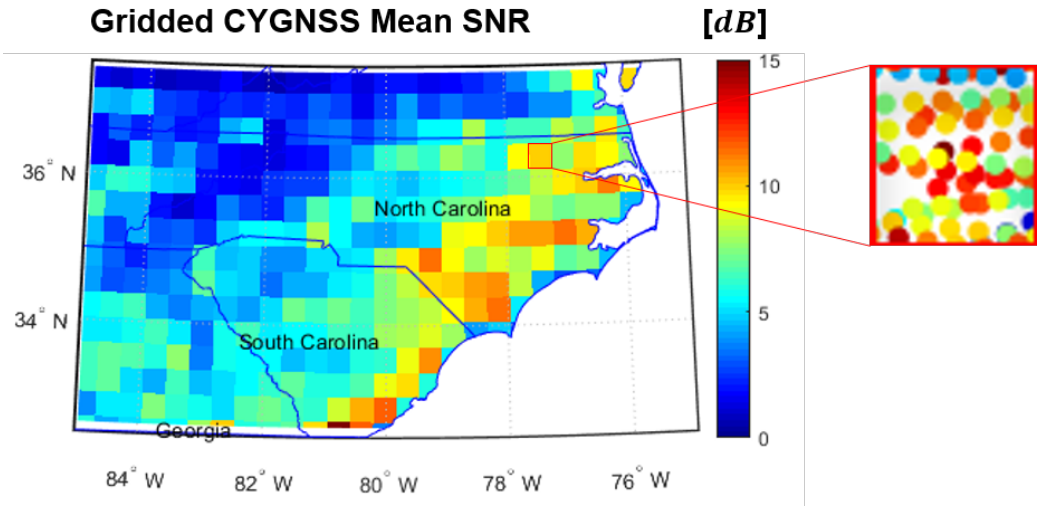

Figure 5.1: Multiple CYGNSS observations (Uncalibrated DDM SNR) that are averaged into the SMAP grids. Data are from 14-18 September 2018 over North Carolina, USA (The landfall of the Hurricane Florence was on 14 September 2018).

5.3 Soil Moisture Retrieval Methodology

We have developed a new, CYGNSS-based SM retrieval methodology that exploits a non-parametric, non-linear machine learning (ML) technique, namely ANN. The decision to use this method is motivated by its following properties and correspondences to the aforementioned requirements as well as challenges of the SM retrieval from CYGNSS observations:

- Non-linear ML algorithms are known for their solid power to solve regression problems where a mix of linear and non-linear dependences exists between parameters [3].
- Such techniques, ANNs in particular, are capable of approximating/learning complex mappings within multi-dimensional parameter spaces with the help of advanced learning algorithms.
- ANNs can, in principle, be trained to approximate any measurable function to any desired degree of accuracy to represent arbitrary inputoutput relations [71]. It should not turn out that the methodology in this study relies on such arbitrary relations. On the contrary, the CYGNSS observables and ancillary data that are major inputs to the regression process are used in order to fulfill the linear/non-linear relations as well as calibration/correction requirements shown in Section 5.2. The property of ANNs makes the use of proxy input features possible for the purpose of fine tuning the overall model performance.
- ANNs are non-parametric models, meaning that the number of parameters that can be input to the retrieval process is flexible, in contrast to the requirement for fixed number of parameters in the parametric models (such as traditional regression models, physical and/or empirical models). This can help advance the CYGNSS-based SM retrieval approaches by introducing the use of additional parameters into the retrieval process. For instance, the CYGNSS trailing-edge slope (TES) can be input into the SM retrieval as a coherency/incoherency indicator in addition to reflectivity, as previously practiced for a study of inundation detection by using another non-parametric learning method [135], instead of dealing with the explicit determination of the coherency.
- The non-parametric nature of ANNs make these models applicable to learn many different kinds of data regardless of their statistical properties. In other words, the retrieval process can integrate data coming from different sources with even poorly-defined (or unknown) probability distributions and relate them well to the parameter of interest [3]. To illustrate, LUT-based SMAP data such as h-parameter (roughness

parameter) and *stem factor* [22] can be incorporated into the CYGNSS SM estimates, combining with *in situ* data such as land cover and NDVI.

- Consequently, such models have to make fewer assumptions about the data distribution, compared to the parametric models. This should not, in turn, mean that the parameters of the CYGNSS-based SM retrieval process have poorly-defined probability distributions. In contrast, it will be demonstrated throughout this section that most of the input parameters coming from CYGNSS observations and ancillary data exhibit well-defined distributions. However, it is a powerful flexibility for ancillary data usage that there is no need to make any assumption about the data distributions.
- The use of such learning algorithms eliminates the need for development of a parametric model that is aimed at explicitly solving the electromagnetic relations and/or relating the *in situ* observations to sensor measurements. This could be beneficial for the SM retrieval from CYGNSS observations to overcome the aforementioned limitations of ancillary data and possibility of too simplistic assumptions.
- ANNs are generally said to be a good balance between accuracy, stability, and computational speed [126].

Throughout this section, insights into the ANN model architecture will be provided first, then detailed information about the data sets that are used in this study will be given, and finally, how the data sets are used in the learning process (training and validation) will be explained.

5.3.1 ANN Model Architecture

In this study, we employ a fully-connected ANN architecture, also known as Multilayer Perceptron (MLP), for the non-linear regression problem, as shown in Figure 5.2. Input features to the learning process are the surface reflectivity, TES, and SP incidence angle from the CYGNSS observations; NDVI and NDVI-derived VWC from MODIS Aqua Surface Reflectance Daily Global 500m data set [166] to represent the vegetation canopy; elevation, terrain slope, and h-parameter values from the CGIAR-CSI SRTM 90m, Version 4 digital elevation model (DEM) database [77] to stand for the surface dynamics. The

selection of these input features were determined after a number investigations on their individual and combined contributions to the estimation performance. The results of these investigations will be provided in Section 5.4 (See Table 5.4). The acquisition of the input features from CYGNSS and ancillary data sets as well as the SM data will be comprehensively explained later in this section.

Reference SM data (in the output layer) are used for optimization of the ANN parameters (minimization of the loss function) in the training stage and assessing the model performance in the validation phase. The proposed model minimizes the loss function, which is defined as the squared error between the model-calculated SM and the reference SM values, over the training data set by running over a predetermined number of iterations. ANN parameters are learned through a stochastic gradient descent solver algorithm, where within each ANN iteration, the model parameters are updated by computing the partial derivatives of the loss function with respect to the ANN parameters (back-propagation) [69, 63]. In other words, the model learns in the training phase the non-linear dependences between the CYGNSS measurements and the reference SM labels with the corresponding ancillary data. Then, the trained model uses these dependences to make future SM estimations for a given set of CYGNSS observations and ancillary data.

In fully-connected ANNs, neurons of one layer are fully interconnected to each other neuron of the adjacent layer. Each layer has a weights-array that can be trained by the forward and backward propagation mechanisms. This array controls the linear strength of the connections to the next layer [69]. Assuming that the number of neurons in i^{th} layer is N_i, the weights-array at the i^{th} layer has a size of $(N_{i+1} \times N_i)$. The inputs-array has a size

of ($8 \times$ M), where 8 is the number of inputs and M is the number of data samples. The result of the matrix multiplication between weights-array and inputs-array of a particular layer is given as input to the next layer. To account for bias in such a linear relation, a trainable bias value is added to the sum at each neuron. The process described so far defines no more than a linear relation in each neuron, and if it was the only operation for the entire ANN, it would only result in a linear regression. The essential part of ANNs that make them powerful to solve non-linear regression problems is the activation function for each neuron. In the literature, a number of different activation functions are used such as Rectified Linear Unit (ReLU), logistic, or tanh activations [117, 80]. These functions are responsible of taking the corresponding bias-added sum as input and transferring it to a new value with the help of the corresponding non-linear relation. This process at each neuron is repeated until the output layer is evaluated, which gives the predicted SM value in this study. The calculation from inputs to the output is named as forward propagation. The network uses the training data and back-propagates the error information by updating the weights and bias in each layer [69] to minimize the defined loss function with the help of stochastic gradient descent algorithm. The entire process makes one iteration of forward and backward propagation. Such iterations are made until the loss function reaches a threshold minimum value or a maximum number of iterations are performed. After the described learning process is finished, the final set of node weights and biases for each layer builds up the trained ANN model for SM predictions. This learned network will produce SM estimates from any new input data parameters through a single forward propagation.

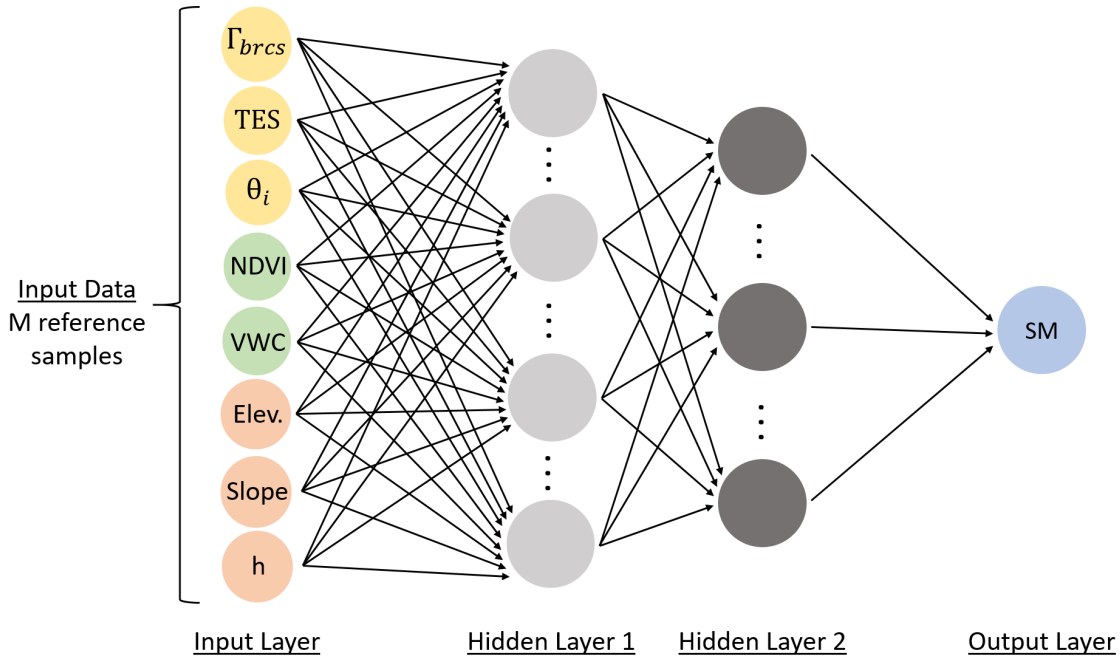

Figure 5.2: Artificial Neural Network with two hidden layers. Input features and the output node are color-coded with respect to separate data sets from which the parameter values were derived. (Γ_{brcs} is the reflectivity calculated from BRCS and will be defined later.)

We have tested several ANN structures, and the ANN parameters that give the best performance out of our investigations are as follows: The input layer has the same number of nodes as the number of used features, which is 8. The output layer has a single node which is the predicted SM values. ANN has two hidden layers in addition to the input and output layers, as shown in Figure 5.2. The non-linear activation function at each layer is chosen to be the Rectified Linear Unit (ReLU) function [117] as it gives the best overall results compared to logistic, or tanh activation functions [80]. Last layer is only a regression layer with no activation function. The Adam solver, which is a first-order gradient-based optimizer for stochastic objective functions [89], is employed for solving optimal weights through loss-function minimization with a learning rate of 0.0001.

5.3.2 Data Sets

This subsection provides details about the acquisition of the SM, CYGNSS, and ancillary data sets as well as their expected contribution to the regression, and the quality control steps to eliminate erroneous data from the analysis.

5.3.2.1 Reference Soil Moisture Data

The present study uses daily SM measurements from *in situ* ISMN sites as reference for the training and validation. The decision to use these data instead of other global SM sources (such as SMAP/SMOS) is built upon three main reasons: **(i)** Comparisons can be made better in a daily basis, compared to 2-3 day revisit time of SMAP and SMOS missions. **(ii)** This study is conducted to investigate high spatial resolution CYGNSS-based SM estimates; however, comparisons with missions like SMAP or SMOS would

require the use of a resolution of roughly 36 km, which under-utilizes CYGNSS's potential. **(iii)** If the learning is done using from SMAP-like satellite observations, the performance of the learning process will be limited by the retrieval performance of that source. For this purpose, a $0.0833^{\circ} \times 0.0833^{\circ}$ lat/lon grid (approximately 9 km $\times$ 9 km around the equator) centering each SM site is considered as a representativeness window where the SM measurements are assumed to be constant. Although this assumption may not always hold as SM can vary much across short distances, it is a necessary assumption in the current state-of-the-science. For instance, the use of SMAP or SMOS missions would require to assume a constant SM value over 36-km regions due to the resolution. Moreover, similar approaches were made in the literature; for example, Dorigo et al. [36] assumed a coarse 50 km window around the ISMN sites. Hereafter, the SM representativeness window around SM sites will be called 9-km-grid in the manuscript.

SM data from 18 ISMN sites throughout North America and Australia are analyzed, considering that there are sufficient number of data samples (CYGNSS observations, reference SM values, and corresponding ancillary data) to input to the ANN model as well as there is enough variability within each parameter (See Figure 5.6). 15 ISMN stations are chosen from Soil Climate Analysis Network (SCAN) sites from the United States of America [146], and the remaining three are from OzNet Hydrological Monitoring Network site from Australia [154]. Detailed information about these SM sites are given in Table 5.2. The locations and International Geosphere-Biosphere Programme (IGBP) land cover types of these ISMN sites are shown in Figure 5.3, which is visualized by using the MODIS/Terra+Aqua Land Cover Type L3 Yearly Global 500m V006 data set [55] via

Google Earth Engine Python API [64]. The selection of these ISMN sites are based on the following reasons: **(i)** The latitudinal coverage of CYGNSS (38° north and 38° south) limits the use of several networks; for example, no station from Europe can be included in this study. **(ii)** Because GNSS-R is sensitive only to the top 5 cm of the soil, the ISMN sites that measure SM at this depth are considered comparable to the CYGNSS estimations [36]. For instance, the COSMOS network would be an alternative network, but their measurements are taken from a varying interval of 0-39 cm. **(iii)** The uniformity of the sensor technology is another constraint. Despite the internal uncertainty of each SM probe, choosing SM networks of the same SM measurement technology would avoid additional biases between the networks. SCAN and OzNet sites are chosen because most of the stations of both networks employ the same instrument (Stevens Water Inc, Hydraprobe). **(iv)** The diversity and temporal coverage of the published data by the networks is also significant as the present analysis needs annual SM data for 2017 and 2018 as well as the temperature measurements for a quality control. For example, the PBO_H2O network was not included in the analysis because their sites provide SM only (Additionally, their measurements are not based on a physical SM probe; instead, their sites are examples of GNSS interferometric reflectometry based SM estimation). **(v)** Most of the ISMN sites in the data set are located on relatively flat (non-mountaneous) surfaces with low-to-moderate vegetation cover (such as croplands, grasslands, savannas) for the sake of limiting the incoherent scattering effects in the analysis.

The SCAN sites provide daily mean SM (measured from top $2.5 - 5$ cm of the soil), air temperature, and precipitation measurements for annual periods. The OzNet sites provide

20-minute SM (for top 0-5 cm of the soil), soil temperature (measured at top 2.5 cm), and precipitation measurements for annual periods. Therefore, their measurements are preprocessed in order to obtain daily averages in this study. Measurements from years 2017 and 2018 for both SCAN and OzNet networks are used in this study since CYGNSS measurements have been made available starting from mid-March, 2017.

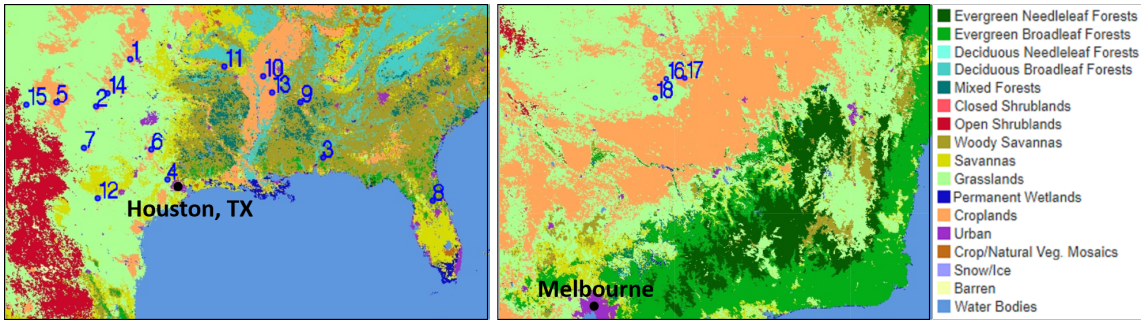

Figure 5.3: ISMN site locations with the IGBP land cover classification in the background. Left: SCAN sites from USA, middle: OzNet sites from Australia, right: IGBP land cover classification color-code. Indices are same with Table 5.2.

5.3.2.2 CYGNSS Data

The CYGNSS observations with a SP location that fall into the 9-km-grid of any of the ISMN sites throughout 2017 and 2018 are included in the analysis. We used the CYGNSS Level 1 v2.1 Science Data Products to obtain the following observables as CYGNSS-representative inputs to the SM retrieval algorithm: **(i)** Reflectivity, **(ii)** SP incidence angle, **(iii)** TES. The definition and acquisition of each input feature is as follows:

Reflectivity is the primary CYGNSS deliverable that must be input to the regression since it is the GNSS-R receivers' observation of the changing SM values and surface con-

Table 5.2: Soil moisture sites that are analyzed in this study. Elevation (in meters) and slope values are calculated as mean values over the 9-km-grid.

ISMN Site	Location	Lat, Lon	Land Cover (IGBP)	Elev.	Slope
1 - Fort Reno (SCAN)	Oklahoma, US	35.55, -98.02	Croplands	420	1.15
2 - Knox City (SCAN)	Texas, US	33.45, -99.87	Croplands, Grasslands	446	0.98
3 - Perdido Riv Farms (SCAN)	Alabama, US	31.11, -87.55	Crop., Savan., Forests	94	1.21
4 - Prairie View (SCAN)	Texas, US	30.08, -95.98	Grass., Savan., Crop.	80	1.12
5 - Reese Center (SCAN)	Texas, US	33.62, -102.04	Croplands, Grasslands	1014	1.14
6 - Riesel (SCAN)	Texas, US	31.48, -96.88	Grasslands, Croplands	161	1.22
7 - San Angelo (SCAN)	Texas, US	31.55, -100.51	Grasslands	600	0.81
8 - Sellers Lake (SCAN)	Florida, US	29.10, -81.63	Forests, Savannas	24	1.87
9 - Starkville (SCAN)	Mississippi, US	33.63, -88.77	Savan., Forests, Grass.	75	1.46
10 - Uapb Marianna (SCAN)	Arkansas, US	34.78, -90.82	Croplands	63	0.84
11 - Uapb Point Remove (SCAN)	Arkansas, US	35.21, -92.92	Croplands, Grasslands	99	1.28
12 - Uvalde (SCAN)	Texas, US	29.22, -99.76	Grasslands, Croplands	285	1.21
13 - Vance (SCAN)	Mississippi, US	34.07, -90.35	Croplands, Forests	47	0.81
14 - Vernon (SCAN)	Texas, US	34.02, -99.25	Grasslands, Croplands	373	1.28
15 - Willow Wells (SCAN)	New Mexico, US	33.53 -103.63	Grasslands	1336	0.97
16 - Yanco-Y7 (OzNet)	Yanco, Australia	-34.85, 146.12	Grasslands, Croplands	128	0.55
17 - Yanco-Y8 (OzNet)	Yanco, Australia	-34.85, 146.41	Croplands, Grasslands	150	0.94
18 - Yanco-Y11 (OzNet)	Yanco, Australia	-35.11, 145.94	Grasslands, Croplands	114	0.75

ditions. The surface reflectivity can be derived from the CYGNSS data products by several ways as described previously in this manuscript and performed by other studies. Four different derivations were investigated in this study: **(i)** we calculated an approximate reflectivity by substituting the DDM SNR (*ddm_snr*) into P_{RL}^{coh} in (5.4) and calibrating for the instrumental and geometric parameters, as done previously [24, 28]. This calculation is called Γ_{snr} in this manuscript. **(ii)** Γ_{peak} is generated similarly to the former approach except that the peak value of the analog power DDM ($power\_analog$) instead of DDM SNR is used for P_{RL}^{coh}. For cases where error level in the DDM noise floor is high (DDM SNR and Γ_{snr} in turn would get erroneous), Γ_{peak} could provide increased correspondence to SM. **(iii)** $brcs$ is used to calculate the reflectivity as shown in (5.5), correcting the incoherency assumption by applying the coherent equation as well as compensating the path loss and 4π term; this reflectivity is called Γ_{brcs} [135]. **(iv)** Γ_{ratio}, was derived by using the ratio of the reflected and direct SNRs (ddm_snr and $direct\_snr$, respectively), which are first calibrated by the range terms, as previously practiced [19]. Separate and combined effects of these reflectivity calculations were investigated in the SM retrieval, and the results reported in Section 5.4 demonstrate that Γ_{brcs} alone has given the highest learning performance. This can be attributed to diverse levels of errors coming from changing calibration parameters in different reflectivity calculations. Therefore, the term "reflectivity" will, hereafter, be used for Γ_{brcs}, unless otherwise stated.

SP incidence angle (in degrees) is used as given in the CYGNSS data. Incidence angle should be taken into account in the CYGNSS-based SM retrieval methods because of two reasons: **(i)** CYGNSS observes the Earth surface over a wide range of incidence angles

spanning from 0° to 70° with a mean of approximately 30° and a standard deviation of roughly 17°, **(ii)** Observed reflectivity values are dependent on the SP incidence angle [2]. Calibrating the reflectivity values for changing incidence angles can be done by two techniques: **(i)** Normalize the reflectivity values at any angles to their corresponding level at 0° by using a curve fit function, **(ii)** Input the incidence angle as a feature to the learning model and let the model capture the angle dependent curve fit for the reflectivity values. We examined both of these approaches, implementing the former by applying as Al-Khaldi et al. considered [2]. Statistical results of the investigation are demonstrated in Section 5.4; however, using the SP incidence angle as an input feature to the ANN model (latter approach) worked slightly better. We attribute the weaker performance of the former to the fact that the curve fit function is based on empirical observations for typical loam soil parameters, which may not be the case for all the ISMN sites in our analysis.

TES is computed as the slope of the trailing edge of the reflectivity delay waveform, as defined by Rodriguez-Alvarez et al. [135]. TES calculation is dependent on the shape of the CYGNSS DDMs and is, therefore, directly related to the coherency/incoherency of the GNSS-R signals. More incoherent mixing through the scattering surface makes TES smaller [135]. Even though this study assumes the dominance of the coherent reflections, we consider the inclusion of TES in the SM retrieval method to be useful for feeding the regression with a coherency/incoherency metric.

5.3.2.3 Ancillary Data

The use of ancillary data as input to the retrieval process is required since the GNSS-R reflectivity is sensitive to not only SM but also other geophysical parameters such as vegetation canopy, topography, surface roughness, and soil texture. It should also be noted that the calculated reflectivity involves the effects from these parameters and are not corrected prior to the retrieval in this study. We used several data sets to represent these parameters as the following input features to the learning model: **(i)** NDVI, **(ii)** VWC, **(iii)** Elevation, **(iv)** Slope, **(v)** h-parameter (Roughness parameter). These input features are computed for every CYGNSS data sample in the analysis and given as input to ANN with the corresponding CYGNSS observables and SM value.

Before getting insight into the ancillary data acquisition, the spatial resolution of this algorithm should be explained. As defined in [135], the semi-major and semi-minor axes of the first Fresnel zone ellipse, where the coherent signals come from, varies between 0.6 km and 0.9 km, as well as 0.6 km and 2.3 km, respectively, for a change of SP incidence angle from 0° to 65°. Depending on the relative orientation of the CYGNSS spacecrafts and GPS transmitters, the first Fresnel zone gets a varying orientation with respect to the along-track direction of the CYGNSS receivers as well. The distance traveled by the SP during the incoherent integration of the GNSS-R signals for one second is roughly 6 km, and it adds an elongation effect to the first Fresnel zone along track direction. In the marginal case where either **(i)** the semi-major, or **(ii)** semi-minor axis aligns with the along-track direction, such an elongation would only affect that axis. As a result, the final size of the CYGNSS footprint, which can no more be an ellipse after the elongation, should, in

marginal case **(i)**, vary between 6.6 km and 8.3 km in the along-track direction whereas the cross-track direction would only vary from 0.6 km to 0.9 km, depending on the incidence angle variations. Therefore, the spatial resolution of CYGNSS would vary between 0.6 km $\times$ 6.6 km and 0.9 km $\times$ 8.3 km. Similarly in the marginal case **(ii)**, the the spatial resolution of CYGNSS would vary between 0.6 km $\times$ 6.6 km and 2.3 km $\times$ 6.9 km. All other possibilities of the footprint orientation on the surface would result in a spatial resolution in between the minimum and maximum of these marginal cases. On the other hand, computations of the Fresnel zone ellipse would have errors depending on the elevation because of the current SP calculation method of CYGNSS that considers Earth as an ellipsoid without topography [28]. Based on these data facts, we considered a $0.04^\circ \times 0.04^\circ$ (approximately 4 km $\times$ 4 km) lat/lon grid cell that centers the SP could be capable of generating the mean terrain statistics (elevation, slope, NDVI, etc.), which in turn was assumed to correspond to the CYGNSS footprint of interest and define the geophysical conditions around the SP location. This grid cell makes an approximate 4 km spatial resolution for SM retrieval since each CYGNSS observation in the analysis is used separately to generate a SM value. Hence, this grid will be called 4-km-grid from now throughout the manuscript. Figure 5.4 shows an example of such grids from the analyzed data set. The descriptions of the NDVI, IGBP land cover classifications, elevation and slope, and inland water body data sets will be given later in this section.

NDVI represents the vegetation cover above the ground as it was previously exploited by the SM missions [86, 124]. Although by definition NDVI ranges from -1 to 1, normally it is positive, and getting values closer to 0 or 1 for sparse or dense vegetation canopies,

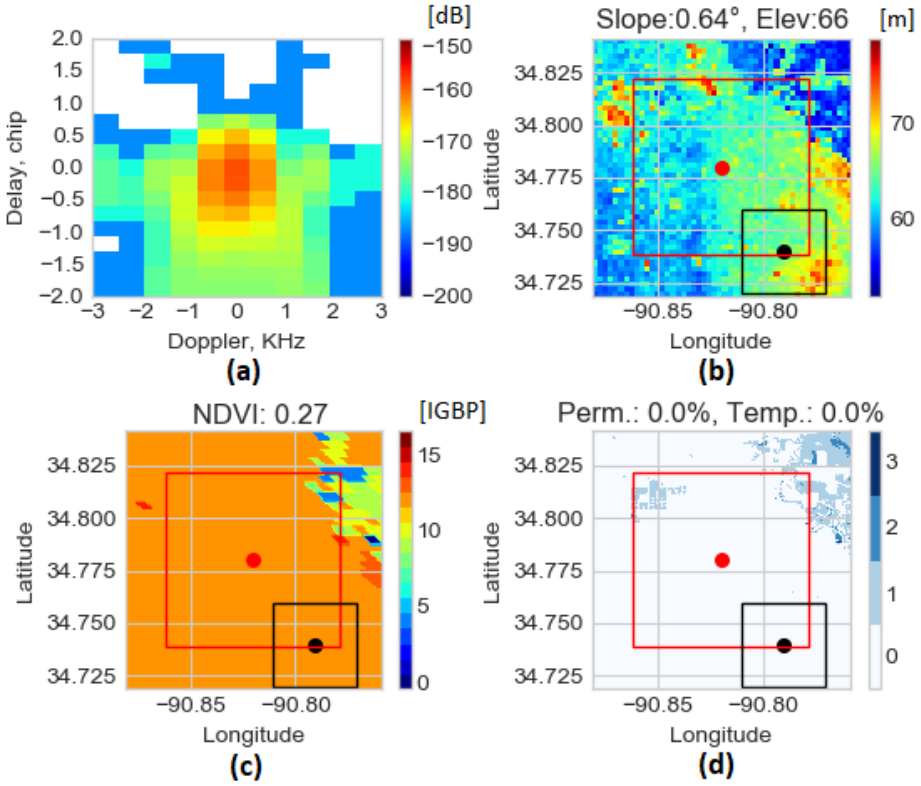

Figure 5.4: Example grid analysis from a CYGNSS observation over the SCAN site Uapb-Marianna on day-of year (DOY) 71, 2018. Red square is the 9-km-grid cell that centers the ISMN site (red dot). Black square is the 4-km-grid cells that center the CYGNSS observation SP (black dot) that could fall into anywhere in the 9-km-grid. **(a)** DDM $power\_analog$. **(b)** Digital elevation model [m]. Mean slope and elevation values calculated from 4-km-grid are also shown. **(c)** IGBP land covers [IGBP]. The color scale for the IGBP land cover is the same with that of Figure 5.3. Mean NDVI calculated from 4-km-grid is also shown. **(d)** Inland water bodies [Pekel indices [128]]. Permanent and temporary water body percentages within 4-km-grid are also shown.

respectively. Two different, dynamic data sets were investigated for NDVI data throughout 2017 and 2018: **(i)** The NOAA Climate Data Record (CDR) of AVHRR Daily NDVI 0.05°, Version 4 [165], **(ii)** MODIS Aqua Surface Reflectance Daily Global 500m data set [166]. The latter has been chosen for its higher spatial resolution. NDVI is calculated from the near-infrared (NIR) and red bands (RED) of the reflectance data as $NDVI = (NIR - RED)/(NIR + RED)$. The mean NDVI for the 4-km-grid of each CYGNSS observation is generated from the MODIS data with the following methodology: The data set is accessed via Google Earth Engine Python API [64] to rapidly and accurately perform analysis for multiple CYGNSS observations at a time. It is possible with the help of very high computing power of the Google Cloud Platform while benefiting the same interface for all different satellite data sets. NDVI data usually suffer from clouds because it is generated by optical instruments such as in the MODIS mission. To deal with this problem, we applied a sliding window averaging over 16 days, whose center is the day of interest (8 days ahead and 7 days to the past). Noting that the 4-km-grid houses 8×8 spatial pixels, we have an $8 \times 8 \times 16$ NDVI cube for each day. To eliminate the ill (cloud-suffered or so) NDVI values, we only considered the width of two standard deviations of the distribution within the NDVI cube to calculate the mean NDVI for that particular grid and day of year [131]. Such an approach would produce close numerical values for adjacent days or short time-series; however, it is the representation of the reality rather than being a problem. More precisely, NDVI experiences quasi-constant trends in daily or weekly periods, and shows stronger dynamics through seasonal changes. Furthermore, spatial variations exist

within the representativeness grid of each ISMN site mostly due to the mix of land covers. Land cover maps for the ISMN sites can be seen in Figure 5.5.

VWC is computed by using the NDVI data via equation (5.10). To the authors' knowledge, the accuracy of this empirical relation for high spatio-temporal CYGNSS resolution is not yet proven. Nevertheless, we decided to use VWC as an input feature in addition to NDVI since it encapsulates land cover information through *stem factor* and temporal memory information through minimum and maximum NDVI values. The *stem factor* information is a land cover-based LUT in the SMAP mission [22]. The *stem factor* value for the 4-km-grid of each CYGNSS observation in this study is calculated as a weighted sum of LUT *stem factor*s based on the land cover percentages in the scene. We performed two different VWC calculations: **(i)** $NDVI_{max}$ and $NDVI_{min}$ are computed from 2017 and 2018 NDVI data. **(ii)** The current NDVI is used in place of $NDVI_{max}$ for the entire data set, and a global constant value of 0.1 was used for $NDVI_{min}$, as suggested in the SMAP's VWC report [22] (Though suggestion for $NDVI_{max}$ was for croplands and grasslands only). The former performed better in the SM retrieval, and we link this to the phenomenon that ISMN sites in this study were not only from croplands and grasslands as the latter method assumes.

Elevation and slope are used to assess their contribution to SM estimation as proxy parameters for the terrain topography, as topography is known to have impacts on the reflectivity [172]. The use of elevation is aimed at helping the regression model learn the impact of the ellipsoid-based CYGNSS SP calculation on the CYGNSS observations, if possible. The slope, on the other hand, is included in the input features as a coherency/incoherency

indicator that could be linked to CYGNSS's TES by ANN. Mean elevation and mean slope for each 4-km-grid are generated from the static CGIAR-CSI SRTM 90m, Version 4 DEM database [77], as similarly done in the literature [135]. DEMs of the ISMN sites can be seen in Figure 5.5.

The *h-parameter* is assumed linearly related to the root-mean-square-height surface roughness (5.7) [124]. Therefore, it is input to the SM retrieval in this study to assess its contribution to account for the surface roughness. *h-parameter* values are listed in a land cover-based LUT in the SMAP mission, similar to *stem factor* [22]. Therefore, we applied a similar calculation for *h-parameter* values over the 4-km-grids as a sum of LUT *h-parameter* values that are weighted by the land cover percentages in the scene.

Distributions and corresponding statistics of the input features can be seen in Figure 5.6. These distributions were analyzed before and after the quality controls were applied. However, the distributions and statistics of the data after application of the quality controls did not drastically change. Since the quality controls are not yet described, only the statistics before the quality controls are given. Figure 5.6 shows that reflectivity follows an almost-perfect Gaussian distribution, so 68% (one-sigma) of the reflectivity values fall in a dynamic range of around 10 dB, whereas 95% (two-sigma) of them are in a dynamic range of roughly 20 dB. This is in parallel to the previously observed and simulated dynamic ranges of around 15 dB over cropland growth seasons [39, 51]. Incidence angle follows a quasi-Gaussian distribution that reflects the variation of the CYGNSS observations, where 95% of the measurements have an incidence angle in a range of approximately $[0, 60]^{\circ}$. The TES distribution is mostly around the mean value of -13.81 with few obser-

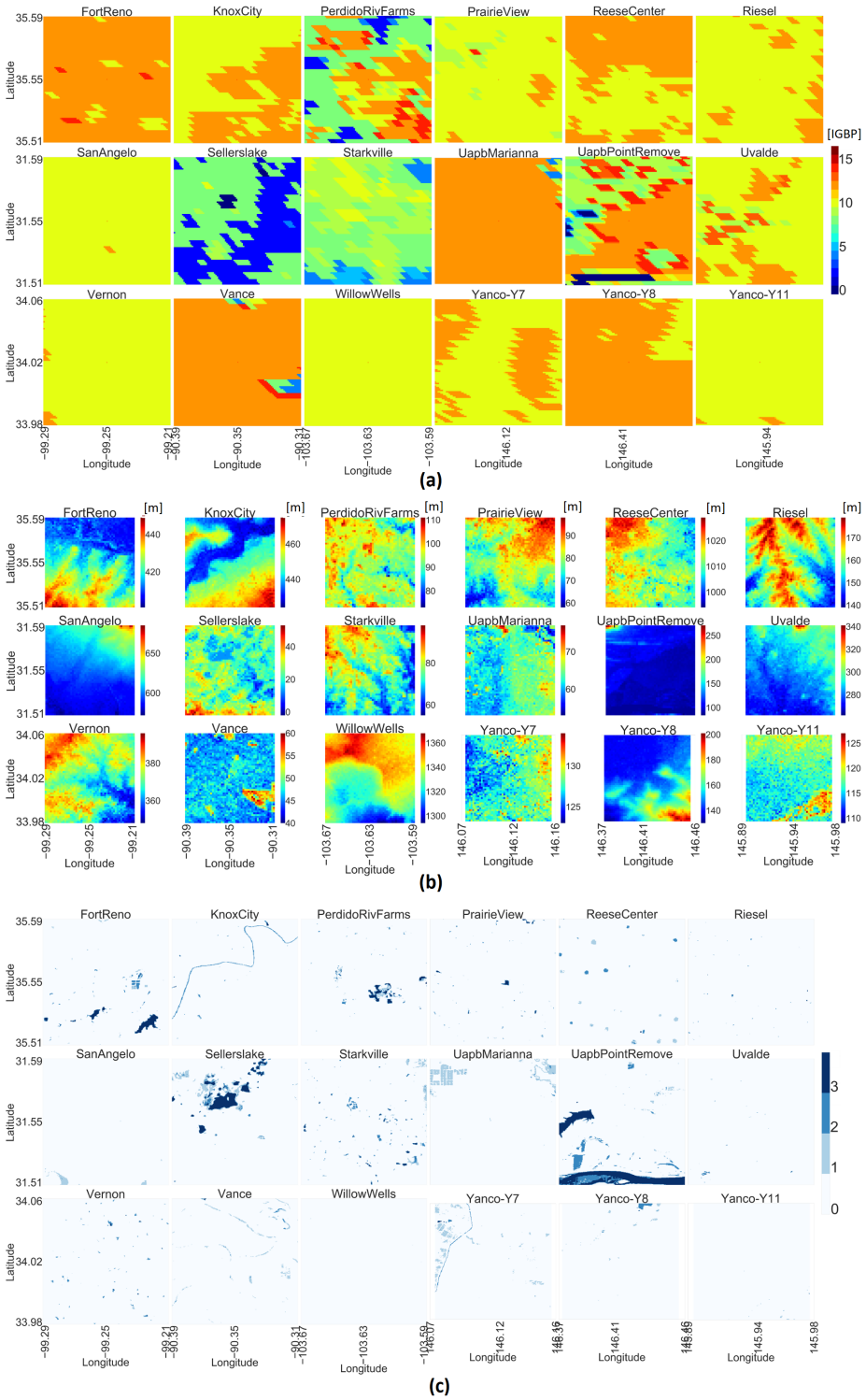

Figure 5.5: Ancillary data visualization for 18 ISMN sites. **(a)** IGBP land cover [IGBP]. The land cover color scale is the same as in Figure 5.3. **(b)** DEM [m]. DEM color scale is generated separately per each ISMN site. **(c)** Inland water bodies [Pekel indices [128]]. In land water body color scale is the same as in Figure 5.4.

vations through larger negative values, which imply that the trailing edge of the reflectivity waveform from peak has a slope of -13.81 on average. The histogram of the elevation data shows the average elevation values around the ISMN sites. The slope distribution indicates that the regions generally have no topographic relief (majority of data under a terrain slope of 2); with a sufficient variation within a slope range of roughly [0.45, 2]. The h-parameter values show no more surprising evidence than that two peaks occur at the SMAP h-values of 0.108 and 0.156 that correspond to croplands and grasslands, respectively [124], and the rest of the data are distributed in between these values. NDVI data prove a good range of variation, where roughly 95% of the NDVI values fall in an approximate range of [0, 0.58] since data seem to follow a nearly perfect Gaussian distribution. VWC, as derived from NDVI by using *stem factor* values from the SMAP-based LUT, distribution seems to be a mix of two Gaussian distributions around mean values (approximately 1.5 and 3.5 kg/m^2). These values likely correspond to grassland and cropland averages, which were the two most common land covers observed in this study.

5.3.2.4 Quality Controls

The quality control of the data sets plays a significant role in the preprocessing of the data for the SM retrieval. We made use of distinct quality control mechanisms for *in situ* SM measurements, CYGNSS observables, and ancillary data. The ideal impact of each quality control step is given Table 5.3, where percent changes in the original dat set is provied as if a particular quality flag is applied alone. In fact, the quality control flags were applied in the left-to-right order in Table 5.3.

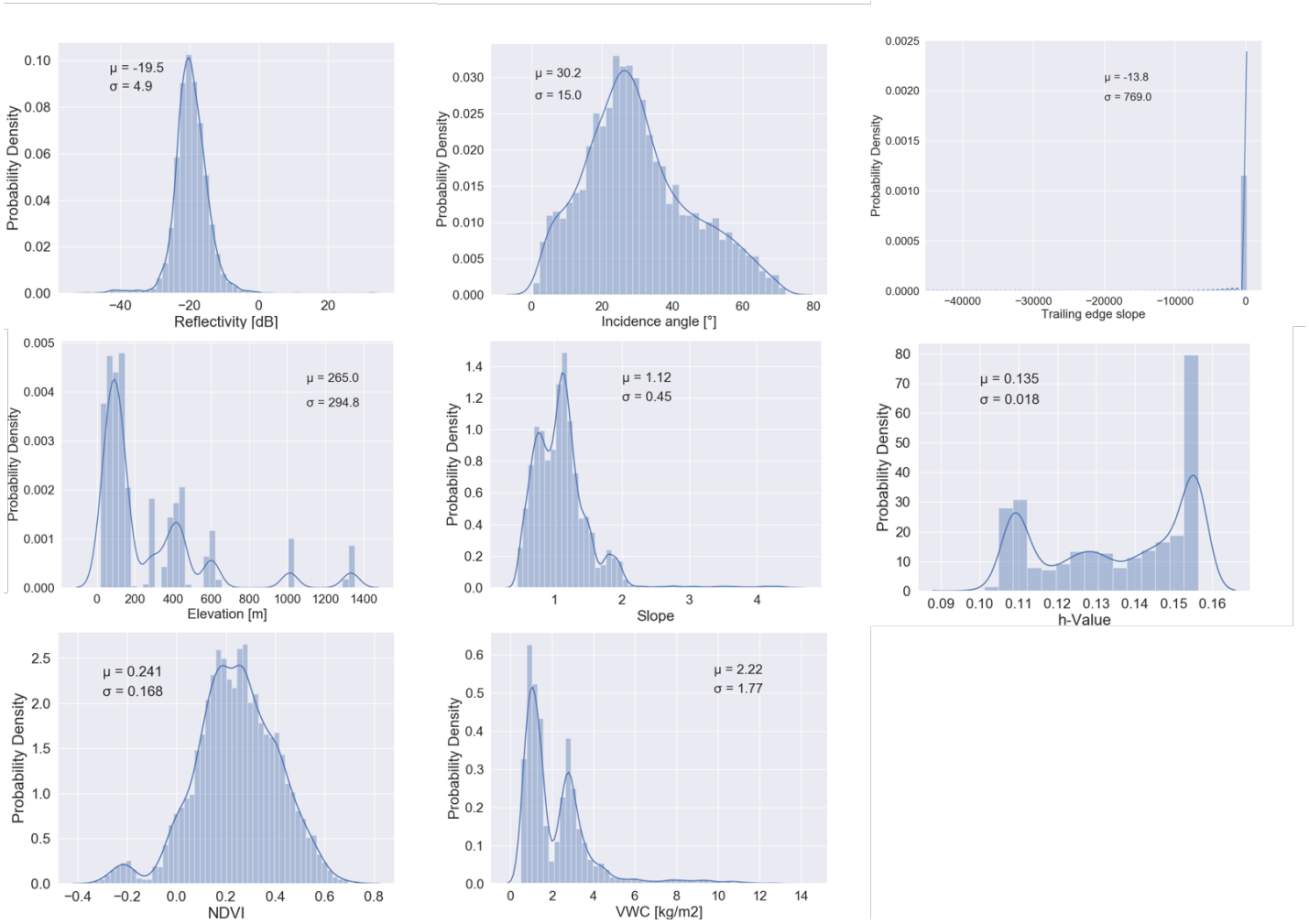

Figure 5.6: Distributions of the input features (before quality control flags applied). Application of quality controls did not result in large variations. Distribution statistics are also given in each plot.

Invalid SM values from the ISMN measurements (such as negative SM or precipitation value) are filtered out of the refrence SM data for both SCAN and OzNet networks. For SCAN sites, SM measurements that correspond to air temperatures below 1° C are excluded from the analysis due to the freezing conditions. The OzNet sites do not experience such conditions as the air temperatures in the region are far away from these ranges annually. CYGNSS observations and ancillary data are not collected for the dates that correspond to invalid and freezing-temperature SM data.

CYGNSS observations require further care to discard the low quality observations from the training and validation. We investigated two different sets of CYGNSS quality flags as Chew et al. [24] and Rodriguez-Alvarez et al. [135] previously applied and observed higher performance with the former. Despite lack of sufficient investigation on the performances of these two quality flag sets, the weaker performance (ubRMSE=0.0557) of the latter might be attributed to the reduced number of data samples used in the training (4027 samples after applying flags from [135]). The quality flags used to filter out the CYGNSS data are as follows: *S-band powered up, Large spacecraft attitude error, Black-body DDM, DDM is test pattern, Low confidence GPS EIRP estimate*. CYGNSS measurements with a negative receiver gain estimation are discarded from the analysis. Table 5.3 shows the percent changes in the data set as if each of the quality flags was applied directly to the original data set. This approach helps demonstrate the true impact of each individual quality control to the data set. The negative receiver gain holds a reasonably big portion (26%) of the CYGNSS observations. Observations with a DDM peak value from outside the range $[7, 10]$ (zero-delay corresponds to the 8^{th} delay bin) are also removed to ensure the error

in the CYGNSS SP location estimation due to the terrain elevation is within a reasonable range [24, 135]. In addition, CYGNSS data points with an incidence angle above 65° are removed due to the poor observation quality, similar to [2].

Inland water bodies have a critical impact on the SM retrieval process because GNSS-R signals get a step reflectivity waveform (sharp increase in the reflectivity) due to the very strong coherency over water surface [23, 135, 24, 142, 2]. Such impacts should be removed prior to the retrieval because they would not reflect SM effects in case the surface water is sufficiently large within the CYGNSS footprint. Being "sufficiently large" is commonly considered as even smaller than the first Fresnel zone [135, 24]. Regarding that the CYGNSS spatial resolution would range from a theoretical minimum of 0.6 km to 8.3 km depending on the incidence angles, and relative orientations of the instruments, we considered a size for the open water bodies that is close to the minimum resolution would work to initiate investigations in this study. Hence, we removed the CYGNSS observations where more than one percent of the 4-km-grid is covered by temporary (seasonal) or permanent surface water. We exploited the JRC Yearly Water Classification History, v1.0 data set (a.k.a. Pekel data set) [128], which is a 30 meter-resolution surface water database. Since the data set is only available from 1984 to the year 2015, we used 2015 data for this study. There are four values (0: No data, 1: Not water, 2: Seasonal-temporary water, 3: Permanent water) for any given pixel. We used the values 2 and 3 to perform water body removal, ignoring when or how long the seasonal water body existed in the year. It is evident from Table 5.3 that around 23% of the entire data set is excluded from the analysis due to inland water bodies. This effect is much higher for particular ISMN sites such as

Sellers Lake (Florida, US), Fort Reno (Oklahoma, US), and Starkville (Mississippi, US). Inland water body maps of the ISMN sites can be seen in Figure 5.5.

5.3.3 Training and Validation

The SM measurement data are provided daily; however, the sub-daily availability of the CYGNSS observations for each ISMN site is not reduced to daily basis in this study. That is to say, for a particular day-of-year around one of the ISMN stations, if there are multiple CYGNSS SPs that fall into the 9-km-grid that centers the site coordinates, all of them are included in the analysis. In such a case, a constant SM value is assumed for all of those multiple CYGNSS observations due to being in an ISMN proximity on the same day. This is considered feasible because the geophysical parameters (such as NDVI, VWC, elevation, slope, and h-value) corresponding to each CYGNSS observation would differ from each other due to the spatial variation, which in turn could explain variations in the CYGNSS observations despite uniform SM values.

After the quality control flags are applied, there are a total of 4808 reference samples (distinct feature vectors) from 18 ISMN stations, spanning from the 77^{th} day-of-year of 2017 (starting date of the publicly available CYGNSS science mission data) to the end of 2018. The training and validation sets are organized with the help of a 10-fold cross validation fashion (N-fold in general [129]) as follows: Training and validation are performed together in a total of 10 iterations. In each iteration, **(i)** 10% of the total data samples are randomly selected and excluded from the training data. **(ii)** The ANN model is trained by using the data samples corresponding to the remaining 90% of the data set. **(iii)** Validation

Table 5.3: The separate, ideal impact of each quality control is shown as percent changes to the original data set: If each quality flag was applied alone to the original data set the percent changes to the original data set would be as given from the sixth to tenth columns (CYG.: CYGNSS quality flags, θ_i: θ_i¡65°, G_r: Positive receiver gain, Peak: DDM peak delay bin rows in [7,10], Water: Removal of inland water bodies). For reference, the initial numbers of data samples before the application of the quality controls are given annually in the second and third columns. Percent losses in the data set due to the actual application of the entire quality controls are also given in the fourth and fifth columns. Overall data set sizes and percent changes are given in the bottom-most row. Year 2017 data for the ISMN sites, Reese Center and Willow Wells, have not been analyzed because these stations are located over 1000 meters, and CYGNSS has an elevation upper limit of 600 meters for the SP calculation algorithm until December 2017 [24]. Year 2018 data for the Riesel station have not been analyzed as well due to large blocks of invalid SM data.

ISMN Site	**Initial #**		**Final (%)**		**CYG.**	θ_i	G_r	**Peak**	**Water**
	2017	**2018**	**2017**	**2018**	**(%)**	**(%)**	**(%)**	**(%)**	**(%)**
Fort Reno	271	275	69.0	66.9	5.2	2.3	30.4	11.4	53.1
Knox City	257	281	58.4	56.2	6.9	2.6	27.9	9.6	37.2
Perdido Riv F.	248	306	55.6	52.9	6.4	2.2	24.6	11.7	38.9
Prairie View	292	289	31.2	31.8	7.3	0.9	24.9	9.6	4.9
Reese Center	—	283	—	29.0	6.4	1.4	21.9	6.7	4.24
Riesel	252	—	33.3	—	6.8	0.4	27.8	12.7	0.0
San Angelo	249	308	48.6	29.3	6.3	0.2	21.7	20.9	5.7
Sellers Lake	246	255	74.0	69.8	7.0	2.3	19.6	12.0	61.9
Starkville	207	272	72.9	74.3	6.7	1.7	26.6	12.9	61.4
Uapb-Marianna	226	267	37.2	32.2	9.1	1.1	26.8	13.1	7.9
Uapb-Point R.	270	306	73.0	75.2	5.1	1.6	28.7	9.6	64.2
Uvalde	236	298	25.4	28.9	5.7	1.4	23.7	10.0	0.0
Vance	226	330	41.6	40.3	7.1	2.8	29.4	19.0	13.4
Vernon	232	294	33.6	28.6	5.9	1.5	26.6	9.8	1.2
Willow Wells	—	292	—	26.4	5.1	0.3	22.6	7.9	1.37
Yanco-Y7	210	287	34.8	41.8	6.2	0.9	26.7	7.7	13.5
Yanco-Y8	215	322	40.4	37.0	7.2	2.0	26.8	8.7	13.3
Yanco-Y11	235	297	30.6	31.7	5.9	2.7	27.6	8.7	0.0
Overall	3872	4962	47.8	43.9	6.5	1.6	25.9	11.2	22.9

of the trained model is performed on the excluded data samples. By this way, the trained model is tested by using "unseen" data in every iteration. **(iv)** The predicted SM values are stored in the prediction pool with their corresponding reference SM values. **(v)** Excluded 10% of the data are turned back into the training data set. The next iteration is performed similarly with the following exception: The random selection of new 10% of the data to be excluded for validation purposes is handled in a way such that samples in this validation set were never chosen into the validation set of any previous iteration. This regulation ensures that the training/validation split method validates the ANN model over the entire data set after 10 iterations are finished. **(vi)** When all 10 iterations are run, the prediction results for the entire data set with corresponding reference SM values are stored. **(vii)** Performance assessment of the entire data set is performed in order to obtain overall and per-ISMN-site statistical performance metrics such as RMSE, unbiased RMSE (ubRMSE), bias, and Pearson's R (correlation coefficient) [43]. The overall training and validation approach is illustrated in Figure 5.7.

5.4 Results

Contributions of individual input features and their combinations to the learning process were first assessed in order to determine the optimal input feature set for SM retrieval. Table 5.4 shows the results of this assessment by using the indices from **(i)** to **(xi)** for changing input combinations. The bottom-most line provides the optimal set of input features that is employed in this study. **(i)**, **(ii)**, **(iii)**, and **(*)** in Table 5.4 were conducted to compare CYGNSS reflectivity calculations, as shown in the top three rows and the bottom-most row.

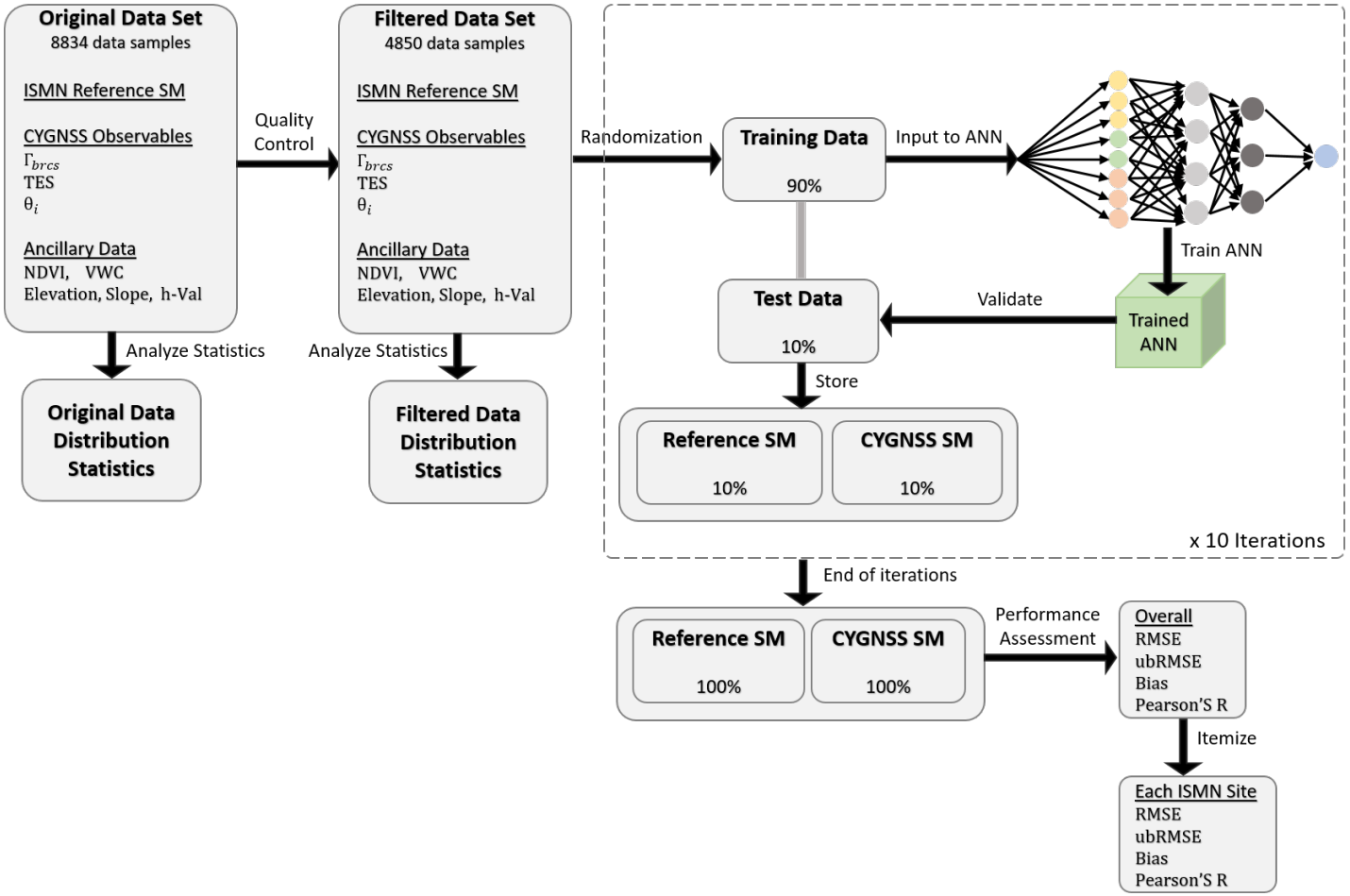

Figure 5.7: Flowchart showing the training and validation of the CYGNSS SM retrieval algorithm.

Γ_{brcs} gives the highest correlation and lowest ubRMSE results. **(iv)** Then, combination of these reflectivity approximations was assessed with two, three, or four of them were used together. All possible combinations were examined; however, only one of them with the closest performance ($\Gamma_{brcs} + \Gamma_{ratio}$) to the optimal one is given in Table 5.4. Although there is a slight difference between two, additional Γ_{ratio} term reduces the overall performance of the optimal combination so that it is not included in the optimal combination. **(v)** Instead of giving the SP incidence angle as an input to the system, the reflectivity values are corrected for incidence and fed to the system without angle information; the model performance was slightly worse than feeding SP incidence angles to the learning. Hence, SP incidence angles were chosen into the optimal performance inputs set. **(vi)** Contribution of TES was examined and found to be significant as its removal increases the ubRMSE and decreases the correlation. TES is also chosen into the input features. **(vii)** and **(viii)** were performed to assess the effects of NDVI and VWC to the learning process. Both have a positive impact on the learning performances, but VWC has a more positive contribution compared to NDVI. This is in parallel with the expectation that VWC involves further information of temporal memory in $NDVI_{min}$ and $NDVI_{max}$,as well as land cover information through $stem\_factor$. Both parameters are included into the input features. **(ix)**, **(x)**, **(xi)** were used to investigate the parameters that are derived from SRTM DEM data set. All of them appear to positively affect the model, but elevation has the highest and h-parameter has the lowest impact on the overall performance. However, all of them are added to the input features set.

Table 5.4: Assessment of the input feature contributions to the learning process. The statistical performance metrics are given for the entire data set. **(*)** The bottom-most row shows the optimal performance, which in turn has the input features set that is used in this study.

Input Features	**RMSE**	**ubRMSE**	**Bias**	**R**
(i) Γ_{snr} + TES + θ_i + NDVI + VWC + Elev. + Slope + h	0.0579	0.0579	0.0006	0.8869
(ii) Γ_{peak} + TES + θ_i + NDVI + VWC + Elev. + Slope + h	0.0635	0.0635	0.0006	0.8619
(iii) Γ_{ratio} + TES + θ_i + NDVI + VWC + Elev. + Slope + h	0.0603	0.0603	0.0007	0.8763
(iv) Γ_{brcs} + Γ_{ratio} + TES + θ_i + NDVI + VWC + Elev. + Slope + h	0.0547	0.0547	0.0010	0.9001
(v) Γ_{brcs} + TES + NDVI + VWC + Elev. + Slope + h	0.0549	0.0549	0.0013	0.9001
(vi) Γ_{brcs} + θ_i + NDVI + VWC + Elev. + Slope + h	0.0596	0.0596	0.0012	0.8920
(vii) Γ_{brcs} + TES + θ_i + VWC + Elev. + Slope + h	0.0554	0.0554	0.0001	0.8970
(viii) Γ_{brcs} + TES + θ_i + NDVI + Elev. + Slope + h	0.0595	0.0595	0.0001	0.8794
(ix) Γ_{brcs} + TES + θ_i + NDVI + VWC + Slope + h	0.0579	0.0579	0.0003	0.8869
(x) Γ_{brcs} + TES + θ_i + NDVI + VWC + Elev. + h	0.0564	0.0564	0.0003	0.8972
(xi) Γ_{brcs} + TES + θ_i + NDVI + VWC + Elev. + Slope	0.0548	0.0548	0.0001	0.8995
(*) Γ_{brcs} + TES + θ_i + NDVI + VWC + Elev. + Slope + h	0.0544	0.0544	0.0002	0.9009

After the determination of the set of optimal input features, validation of the method with this set was performed. Figure 5.8 shows the scatter plots of the ISMN-measured and CYGNSS-retrieved SM values in conjunction with the RMSE, ubRMSE, bias, and Pearson's R values for the entire data set, as well as years 2017 and 2018. In addition, per-site and overall performance statistics of the SM estimation results for the entire data set for both years are shown in Table 5.5. The entire data set, spanning both years, has a Pearson's R value of 0.9009, which is an indicator of high overall agreement between the CYGNSS-based SM predictions and the reference SM data. Data over either year show a high correspondence, too, with Pearson's R values of 0.8532 and 0.9238 for year 2017 and 2018, respectively. These high levels of correlation demonstrate that the presented CYGNSS-based SM retrieval algorithm were successful in capturing the overall trends in approximately five thousand CYGNSS data samples. The ubRMSE values of 0.0575, 0.0520, and 0.0544 cm^3/cm^3 for 2017, 2018, and entire data set, respectively, are obtained. Keeping in mind that both the science mission requirements for the SMAP mission over its calibration/validation sites were ubRMSE values no higher than 0.04 cm^3/cm^3 [124] and CYGNSS land application studies were targeted at an ubRMSE of 0.05 cm^3/cm^3, our algorithm seems capable of generating close values to these levels at least for the current data set and input features of interest. It should be also noted here that, despite not being shown here, a previous version of this analysis was conducted with a subset of the data, input features, and quality controls; the overall performance (overall ubRMSE of 0.0594 cm^3/cm^3 and Pearson's R of 0.6604) was poorer than the ones reported here although six of the ISMN sites with almost the best performances of this study were used. These

observations are of high importance for the ultimate CYGNSS-based, or GNSS-R based in general, SM retrieval studies since the algorithm presented here has the potential to provide increased estimation performances as new satellite observation data and relevant input features are added to the training.

Considering the Pearson's correlation coefficient ranges of [0, 0.3333), [0.3333, 0.6667), and [0.6667, 1] as low, moderate, or high correlations, $3/4$ of the stations have moderate as well as $1/4$ of them have high correlation results through 2017. For 2018, roughly $1/9$ of the sites have low, $3/9$ of them have moderate, and the remaining $5/9$ have high correlation levels. An interesting outcome of the correlation assessment is that even though the highest Pearson's R value for an ISMN site for 2017 is 0.8077 (Uvalde) and 2018 is 0.8909 (Uapb-Marianna), the overall Pearson's R value for each year (0.8532 and 0.9238, respectively) exceeds these maxima. Moreover, the overall correlation coefficient for the entire data set (0.9009) is higher than these maxima as well. This is valuable especially for potential future application of the algorithm to the global scenarios, indicating that it is powerful to generalize the non-linear regression for the entire data despite some poor performances on specific sites. Similar to the correlation, it can be reasonable to consider ubRMSE values above 0.0650 as low-performance, those in the range [0.0500, 0.0650] as moderate-performance, and values below 0.0500 as high-performance. In this case, the algorithm produces low-performance for $1/8$ of the sites, moderate-performance for $3/8$, and high-performance for $4/8$ of those stations for 2017. For 2018, it predicts SM with a low-performance for approximately $1/9$ of the ISMN sites, moderate-performance for $4/9$ of them, and high-performance for another $4/9$. It is worth noting that ubRMSE alone

would be deceptive as some ISMN sites would have very low trends of SM values in yearly average. However, combined with the correlation performances, SM estimations of our algorithm seem to be in a good agreement with mean SM levels of most of the stations. On the other hand, the model performs poorly on particular stations such that the predicted SM values cannot correlate strongly to the reference SM data. For instance, Starkville, Sellers Lake, and Uapb-Point Remove are such SCAN sites where the model's Pearson's R follow low-to-moderate values for both years. These stations have a common feature that more than 60 percent of the observation data are subject to exclusion from analysis due to the existence of inland water bodies. Therefore, the removal process of the inland water bodies and/or the accuracy of the 4-km-grid might require further investigation in the future. San Angelo has another interesting result that Pearson' R for 2017 is much lower than that of 2018. Although this SCAN site is located an elevation of approximately 600 m (which is equal to CYGNSS's 600 m threshold for year 2017 as mentioned previously), we included it in the analysis for both years. Nevertheless, it appears that even an elevation that is equal to the altitude threshold of the CYGNSS SP calculation algorithm for 2017 would be erroneous.

Figure 5.9 and Figure 5.10 show per-site comparisons for selected sites between the CYGNSS-based SM estimations and daily ISMN measurements for years 2017 and 2018, respectively. These figures show both time-series to visualize the annual trends and scatter plots to illustrate the correlation between the reference and predicted SM values. First of all, the ability of the algorithm to generate sub-daily SM predictions for multiple CYGNSS observations on a day can be observed on these figures. For instance, the figure grid-

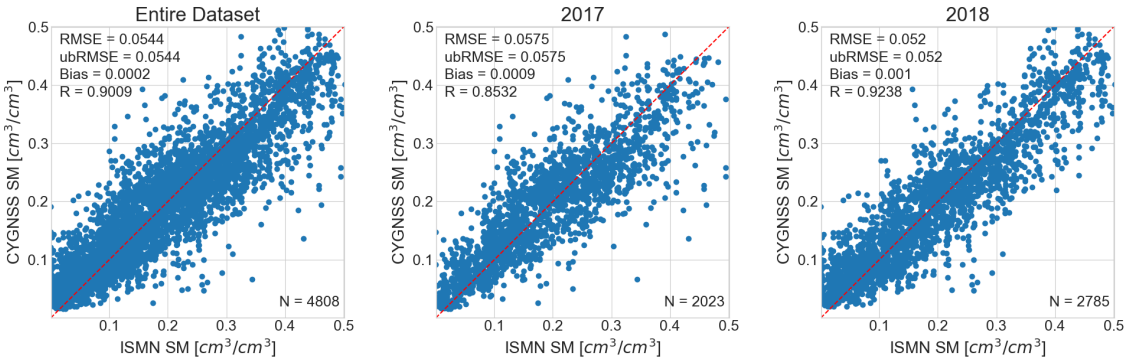

Figure 5.8: Scatter plots of SM retrievals for the entire data set, year 2017, and year 2018. RMSE, ubRMSE, and bias are given in $[cm^3/cm^3]$. Pearson's R is also provided. N is the number of data points.

line that corresponds to October on the top-first plot (Uvalde) in Figure 5.9 shows well that there were two CYGNSS observations that fell into the ISMN site representativeness window on that particular day-of-year, and the algorithm was able to generate two different SM predictions that are so close to the reference SM value. Figure 5.9 and Figure 5.10 are prepared in a way that the top four ISMN sites are selected from among those where the SM retrieval algorithm performed the best, the next one is picked from those with decent results, and the bottom-most ISMN site is chosen from a set of stations where the algorithm has the poorest performance, with respect to both the correlation and retrieval errors. For example, Uapb-Point Remove was selected for the bottom-most plot of Figure 5.10 while Yanco-Y8 is the station with the lowest Pearson's R value (0.2529). This is because even though Uapb-Point Remove has a close Pearson'R to Yanco-Y8, it has a much higher ubRMSE (0.0558) for SM estimations. Similarly, Vance site has a high ubRMSE of 0.0744

for 2018, but it has a high Pearson's R of 0.8489 at the same time. Hence, it was chosen for the bottom-most, too.

5.5 Discussion

The proposed learning-based SM retrieval methodology generates promising overall performances with comparable accuracy to the reference SM data, demonstrating that it could be generalized for the global SM estimation. For this purpose, the current trained model can be used over global terrains with close statistical distributions of the bio-geophysical parameters (vegetation cover, topography, and surface roughness) in the present analysis. Alternatively, the model can be further trained over a much larger and globally-representative data set with additional CYGNSS observations, terrain characteristics, land cover classifications, as well as possible new input features. As more data samples are added to the training, the learning performance is expected to improve. In any case, a future work will be conducted for comparison of the proposed method to global SM sources (such as SMAP). A potential limitation in a global comparison scenario is the lack of reference SM data at high spatio-temporal resolutions. This was the main motivation of this study to employ ISMN sites with a 9-km-grid of SM representativeness. Assuming constant SM over such a grid is another source of limitation in this study, which is inevitable for the current state-of-the-science. The reference SM data, either ISMN sites in this study or SMAP-like global observatory data, are not the ground truth and have their own internal errors. Moreover, ISMN measurements have a quite different resolution (point-scale) than CYGNSS observations (distributed scattering).

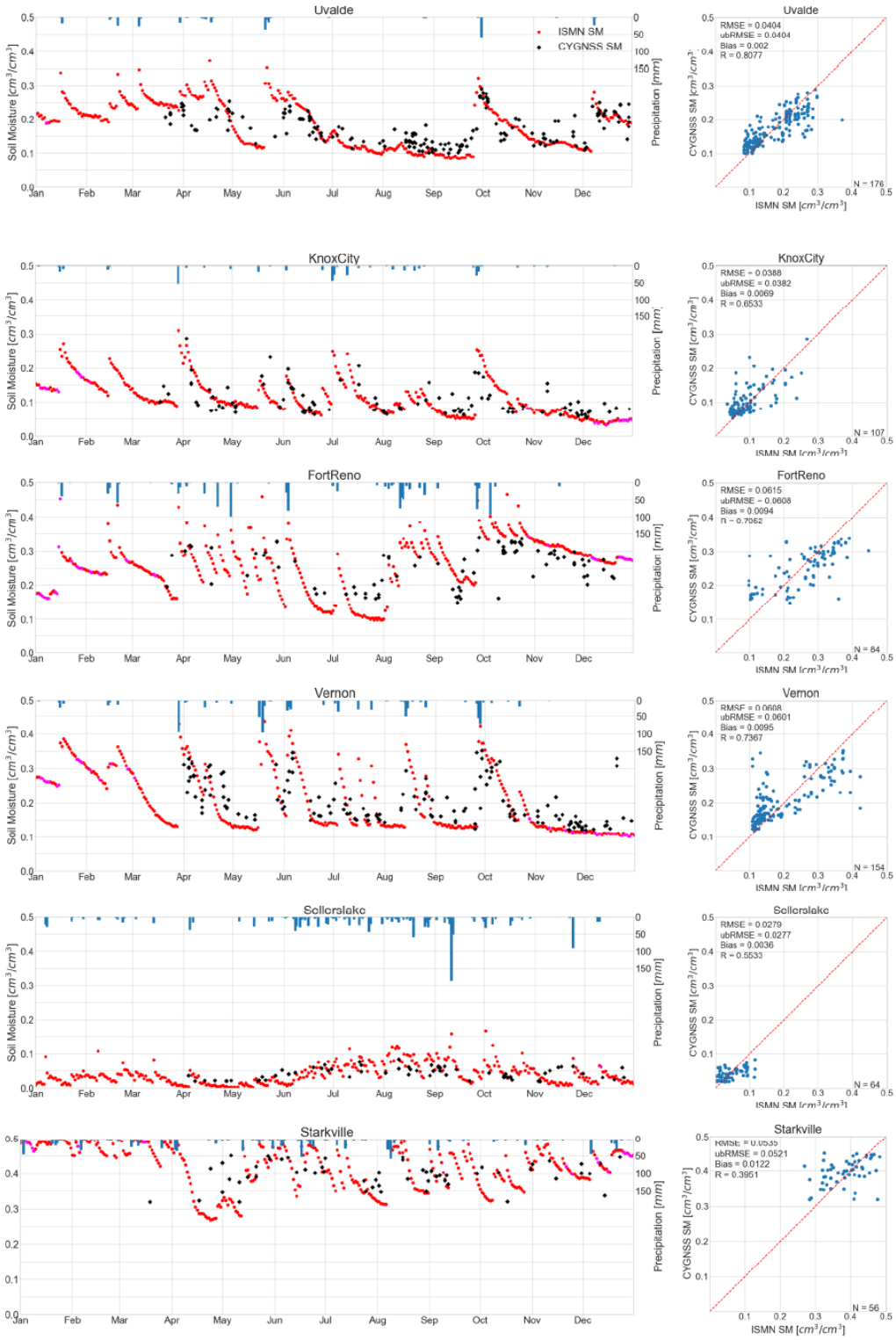

Figure 5.9: SM retrievals for year 2017. *In situ* SM data are color-coded such that magenta and blue circles correspond to days with freeze temperatures and invalid data, respectively. Scatter plots are also given.

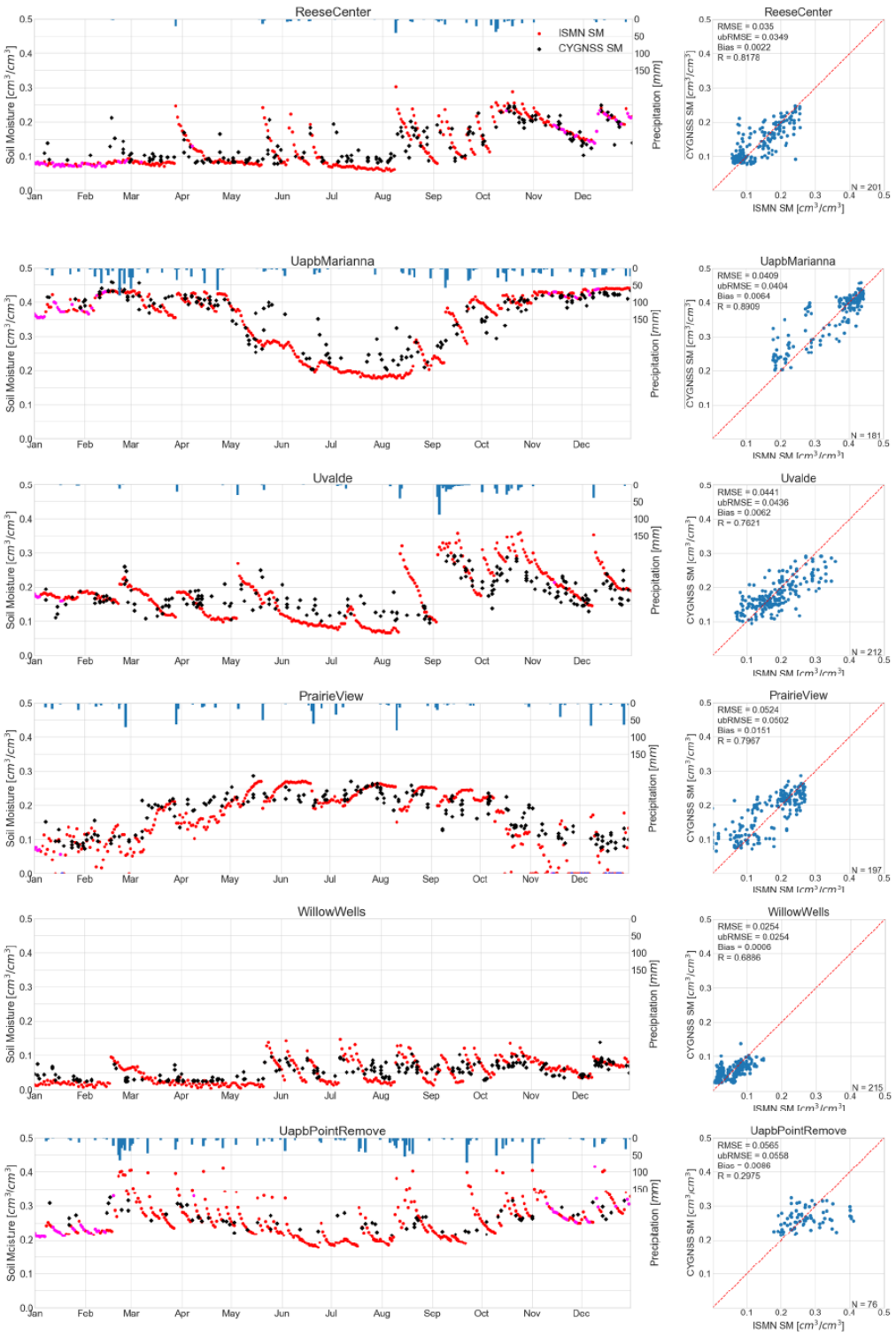

Figure 5.10: SM retrievals for year 2018. *In situ* SM data are color-coded such that magenta and blue circles correspond to days with freeze temperatures and invalid data, respectively. Scatter plots are also given.

Table 5.5: SM retrieval algorithm performance statistics. RMSE, ubRMSE, and bias are expressed in $[cm^3/cm^3]$. Pearson's R is also provided.

	2017				2018			
ISMN Site	**RMSE**	**ubRMSE**	**Bias**	**R**	**RMSE**	**ubRMSE**	**Bias**	**R**
Fort Reno	0.0615	0.0608	0.0094	0.7052	0.0521	0.0520	0.0013	0.8019
Knox City	0.0388	0.0382	0.0069	0.6533	0.0503	0.0492	0.0108	0.7640
Perdido Riv Farms	0.0493	0.0487	0.0080	0.5986	0.0521	0.0513	0.0091	0.4417
Prairie View	0.0465	0.0455	0.0099	0.5897	0.0524	0.0502	0.0151	0.7967
Reese Center	-	-	-	-	0.0350	0.0349	0.0022	0.8178
Riesel	0.0947	0.0918	0.0236	0.6706	-	-	-	-
San Angelo	0.0615	0.0606	0.0103	0.3825	0.0705	0.0687	0.0158	0.6290
Sellers Lake	0.0279	0.0277	0.0036	0.5533	0.0252	0.0246	0.0059	0.4171
Starkville	0.0535	0.0521	0.0122	0.3951	0.0621	0.0614	0.0415	0.3609
Uapb-Marianna	0.0522	0.0514	0.0088	0.5994	0.0409	0.0404	0.0064	0.8909
Uapb-Point Remove	0.0442	0.0363	0.0253	0.3350	0.0565	0.0558	0.0086	0.2975
Uvalde	0.0404	0.0404	0.0019	0.8077	0.0441	0.0436	0.0062	0.7621
Vance	0.0789	0.0776	0.0142	0.5764	0.0750	0.0744	0.0095	0.8489
Willow Wells	-	-	-	-	0.0254	0.0254	0.0006	0.6886
Vernon	0.0608	0.0601	0.0095	0.7367	0.0567	0.0565	0.0056	0.8506
Yanco-Y7	0.0417	0.0411	0.0074	0.5645	0.0358	0.0354	0.0057	0.7348
Yanco-Y8	0.0453	0.0444	0.0089	0.4662	0.0565	0.0553	0.0113	0.2529
Yanco-Y11	0.0604	0.0602	0.0041	0.4410	0.0571	0.0571	0.0008	0.6170
Overall	0.0575	0.0575	0.0009	0.8532	0.0520	0.0520	0.0010	0.9238

In addition to extending the use of the current method, a future work could be conducted to investigate several non-parametric, non-linear machine learning algorithms as well as the optimization of the current ANN model to obtain its full potential. Despite a number of actions to get the best performance out of the learning model (such as preprocessing of the data with the quality controls, assessment of the input feature contributions, and use of derived proxy parameters), the scope of this study is not an in-depth analysis of the learning methods.

The present method has many sources of constraints and uncertainties, some of which have been addressed and dealt with to a degree in this study, and all could be investigated further in future work. These can be explained as follows:

CYGNSS observations and data products involve a number of uncertainties that might have affected the results of this study. During this study, multiple reflectivity approximations from CYGNSS data products were used as model inputs. The learning performance of the model varied greatly with these different reflectivity approximations. This indicates a changing level of uncertainty throughout the CYGNSS dataset. CYGNSS parameter uncertainties can be described as follows: **(i)** SP calculation with the geoid assumption was originally developed and works well for ocean [62]; however, it can generate large offsets for SP locations over land as the elevation and topographic relief get higher. Since this problem is said to be resolved with the CYGNSS Level 1 Science Data, Version 3.0 in a near future, we applied a number of strategies that did not involve explicitly solving the actual SP location: With the help of the coherency assumption, we used the peak *brcs* value to calculate Γ_{brcs}, which should reflect the actual coherent reflections; we filtered

out *brcs* DDM peak delay rows that fall outside a range of $[7, 10]$; we also incorporated CYGNSS-derived TES, elevation, and slope as proxy coherency parameters to the ANN model. In addition, analyzing terrains with relatively low topographic relief (slopes up to 2) and elevations no more than roughly 1300 meters in this study limits the effects of this uncertainty; however, these parameters would get higher values in a global application. **(ii)** As previously discussed, the CYGNSS data products, which are determined with the use of internal and external parameters (such as GPS EIRP and receiver gain, as well as bistatic ranges due to topography and erroneous SP calculation) would introduce an added layer of uncertainties to the input data of the learning process. A subset of these could be partially corrected by future effort; for instance, the bistatic ranges would be obtained more accurately by a corrected SP calculation strategy with the help of DEM data. **(iii)** The present method assumes the dominance of coherent reflections over flat and relatively smooth lands covered with non-heavy vegetation canopy, as followed by several previous studies [142, 178, 24, 28]. Indeed, high variability of the topography and vegetation covers makes it impossible to consider pure coherent or incoherent regimes for global applications. In addition, changes in the observation geometry (such as increasing incidence angles) would introduce increasing incoherent scattering that might be comparable to the coherent reflections. The attempt of this study to introduce coherency/incoherency indicators to the learning process would be improved further by incorporating additional CYGNSS observations (such as the entire DDM) to provide and improved coherency detection. It should be also noted that the use of such parameters would necessitate a trade-off between increased accuracy and decreased spatial resolution.

Ancillary data are required since CYGNSS measurements are dependent on several bio/geophysical parameters in conjunction with SM, but data sources are not perfect. NDVI is used to account for vegetation attenuation because there is no global data set of another parameter on which the reflectivity is more dependent on (such as vegetation optical depth or VWC). NDVI is indeed a metric of vegetation "greenness" and does not fully correlate to the attenuation. Moreover, it is derived from optical imaging instruments, meaning that it is vulnerable to the atmospheric or illumination effects such as clouds or night. Such issues can only be diminished to some degree as this study performs such that a sliding window averaging can be employed. NDVI-derived VWC data and h-parameter have even additional biases as we employ land cover-based LUT values for *stem factor*. On the other hand, all of these parameters prove increased accuracy to the algorithm outcomes. Improved acquisition of such ancillary parameters as well as involvement of new input features (such as several vegetation indices) would be of interest in the future.

Internal steps and decisions as well as assumptions and simplifications of the proposed methodology might have led to issues that are not clear for the time as well. To illustrate, the use of 4-km-grid for averaging the terrain parameters (such as NDVI, elevation, and slope) around the CYGNSS SP might be too simplistic; nonetheless, it relies for now on the present uncertainty of the CYGNSS footprint over land, and it appears to be working. Future work would examine different sizes of such a grid as well as introducing a new grid that is computed accordingly with the along-track direction of the SP.

Although we perform quality controls such as applications of CYGNSS quality control flags, removal of measurements with freezing weather conditions, and exclusion of

inland water body-exposed observations, modification of the current method (Such as a different percentage threshold for inland water bodies) and/or addition of further quality flags would be investigated. However, it is evident from the data statistics before and after the quality controls of the present study that roughly half of the data samples are thrown away. This implies that further quality controls would result in a reduced size of training data and would require the addition of new observations. In addition, the current quality controls such as removal of inland water bodies above one percent might not be sufficient to overcome erroneous observations.

5.6 Conclusion

To the authors' knowledge, this paper is the first to provide a physics-aware, learning-based, CYGNSS SM retrieval algorithm that uses several input features from ancillary data sets. The theoretical background of ideal CYGNSS-based SM retrieval, challenges with the use of CYGNSS observables, and the applicability of the proposed method as well as input features are provided. The acquisition of the data sets are described in detail, and their distribution statistics are given. Quality controls for preprocessing of the data are proposed. The approach has been applied to reference SM data collected from 18 ISMN stations through North America and Australia by using a randomly cross-validated train/test split fashion, and encouraging results have been reported. The trained model has an overall ubRMSE of 0.0544 cm^3/cm^3 and Pearson's R of 0.9009 on 4808 unseen data samples, which indicates a potential to be applied on global data sets. Individual contributions from input features have been also investigated. The use of *brcs*-derived

reflectivity, TES, incidence angle, NDVI, VWC, elevation, slope, and h-parameter appears to significantly increase the model performance.

The major contributions of the study to the GNSS-R based SM retrieval efforts in general are as follows:

- This retrieval process is an intrinsically non-linear regression problem and the proposed ANN model demonstrates potential to overcome this problem with its generalization performance.
- ANN is a non-parametric learning algorithm, and it, in turn, removes the need for development of a parametric, explicit solution model (physical, regression, and/or empirical). Therefore, it enables using arbitrary number of input parameters.
- Once trained on reference samples, the model does not require SM information from other sources (such as SMAP or SMOS) for SM estimation. This allows he learning model to be trained with any SM source (such as the ISMN sites in this study, or some other point-scale or global-scale SM sources) and perform SM predictions independently from other SM providers.
- The algorithm has the potential to provide sub-daily SM estimations as it is applicable to every single CYGNSS observation without a need for spatial gridding or temporal averaging. In other words, it is capable of making a SM prediction by using a single CYGNSS observation supported by the other input features.
- The reported overall performances as well as terain-specific accuracy could be improved by including new data samples into the training as the CYGNSS mission continues to orbit and collect new land data in coming years.
- The proposed strategy can help determination of the calibration/validation sites for CYGNSS-based SM retrieval studies in order to learn the global bio/geophysical dynamics (land covers, topography).

CHAPTER 6

CONCLUSION

This chapter provides the summary of the research as well as future work that can be conducted as the proposed method here is still in its emerging phase and needs to be improved in many ways.

6.1 Summary of Research

This dissertation proposed a non-parametric, non-linear learning algorithm for CYGNSS-based SM estimations by using a number of CYGNSS observables and ancillary data sets as inputs. The proposed ANN model provides a novel non-linear solution to intrinsically non-linear regression problem of the CYGNSS-based SM retrieval. The method is capable of generating sub-daily SM products for a 4-km-representativeness grid of each CYGNSS observation. The outputs of the learning model demonstrates an overall ubRMSE of 0.0544 cm^3/cm^3 and Pearson's correlation coefficient of 0.9009 on approximately five thousand CYGNSS data samples. The nature of the ANN model used has a potential to provide better learning performances by new data samples are added into the learning process as CYGNSS spacecrafts continues measurements. From this perspective, the algorithm shows potential to provide closer accuracies to the operational thresholds.

The following major efforts have been made throughout the Ph.D. research for determination of the proposed methodology:

(i) The requirements and challenges for solving the bistatic radar equations to obtain SM values from CYGNSS measurements have been addressed through the development of the SCoBi model and its simulator framework, conducting simulated analyses with the help of SCoBi, as well as reviewing the literature.

(ii) Dynamic ranges of the joint impacts of various bio/geophysical parameters and varying sensitivity of the CYGNSS receivers to these parameters have been taken into account by the retrieval algorithm. The dynamic ranges of the GNSS-R observables corresponding to each of these impacts were modeled by our simulated GNSS-R study over the entire growth season of a virtual corn crop field [51]. Use of ancillary data in the proposed SM retrieval method relies on representing the joint impacts of the changing parameters in the learning process. In addition, the non-linear nature of the proposed ANN model is considered efficient to overcome solving this non-linear regression problem.

(iii) High spatio-temporal resolution of the CYGNSS observations have a potential to provide SM estimations within finer grids than the current state-of-the-science. Employment of the 4 km-grid around each CYGNSS observation as well as the use of sub-daily CYGNSS measurements in the learning process are for investigating the feasibility of such high spatio-temporal resolutions. Additionally, 9 km-grid is employed around each point-scale SM source (ISMN stations) as a representativeness window for the same reason.

6.2 Future Work

Future work is required to validate the current trained model's performance over global CYGNSS observations. It can be handled in two different approaches: **(i)** The current trained model can be directly used over global terrains. The ancillary data set distributions from the currrent method can be used as determination criteria for whether to include terrains in the global analysis in the estimation process. The model can be run over the CYGNSS observations over those terrains included to perform SM estimations. **(ii)** The training data set can be extended in order to represent a larger subset of the global land covers, and a new model can be trained over this larger data set, then it can be examined over a global test data set that is never seen by the model. The prediction performance of the either approach can be compared qualitatively (i.e. by visualizations) and quantitatively (i.e. by providing performance statistics) to a global SM data source such as NASA's SMAP or ESA's SMOS. The quantitative comparison of the algorithm's SM estimations would require a further effort such that multiple SM predictions corresponding into a single coarse-resolution SM product of either mission should be carefully tackled.

The parameters of the ANN model employed in this study have been optimized to obtain the optimal performance; however, it would require tuning as new data samples or new input features are added into the learning process. Other non-parametric, non-linear learning algorithms would be of investigation as well.

A future work can be conducted to determine the error budget of the method that is resulted from the aforementioned uncertainties in Chapter 6.

The current method assumes the dominance of coherent reflections for the case that the surface is flat and smooth with non-heavy vegetation over it. This was based on out simulated analysis findings from SCoBi modeling studies as well as similar literature efforts. However, a space-borne incoherent modeling study can be conducted as a future add-on to the SCoBi model. On the other hand, it would be one of the hardest future extensions of this dissertation because modeling the surface and vegetation scatterers over large areas (i.e. tens of kilometers) requires efficient modeling techniques as well as high computing power.

REFERENCES

[1] "2017 InVEST Awards: Projects awarded funding under the In-Space Validation of Earth Sciences Technology (InVEST) Program 2017 ROSES A.49 Solicitation NNH17ZDA001N Research Opportunities in Space and Earth Sciences,", NASA Earth Science Technology Office, 2018.

[2] M. M. Al-Khaldi, J. T. Johnson, A. J. O'Brien, A. Balenzano, and F. Mattia, "Time-Series Retrieval of Soil Moisture Using CYGNSS," *IEEE Transactions on Geoscience and Remote Sensing*, 2019.

[3] I. Ali, F. Greifeneder, J. Stamenkovic, M. Neumann, and C. Notarnicola, "Review of machine learning approaches for biomass and soil moisture retrievals from remote sensing data," *Remote Sensing*, vol. 7, no. 12, 2015, pp. 16398–16421.

[4] A. A. Arroyo, A. Camps, A. Aguasca, G. F. Forte, A. Monerris, C. Rdiger, J. P. Walker, H. Park, D. Pascual, and R. Onrubia, "Dual-polarization GNSS-R interference pattern technique for soil moisture mapping," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 7, no. 5, 2014, pp. 1533–1544.

[5] A. A. Arroyo, A. Camps, A. Monerris, C. Rudiger, J. P. Walker, and R. Onrubia, "On the Correlation Between GNSS-R Reflectivity and L-Band Microwave Radiometry," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 9, no. 12, 2016, pp. 5862–5879.

[6] M. Aubert, N. Baghdadi, M. Zribi, A. Douaoui, C. Loumagne, F. Baup, M. El Hajj, and S. Garrigues, "Analysis of TerraSAR-X data sensitivity to bare soil moisture, roughness, composition and soil crust," *Remote Sensing of Environment*, vol. 115, no. 8, 2011, pp. 1801–1810.

[7] B. Barrett, E. Dwyer, and P. Whelan, "Soil moisture retrieval from active spaceborne microwave observations: An evaluation of current techniques," *Remote sensing*, vol. 1, no. 3, 2009, pp. 210–242.

[8] W.-M. Boerner, W.-L. Yan, A.-Q. Xi, and Y. Yamaguchi, "Basic concepts of radar polarimetry," *Direct and Inverse Methods in Radar Polarimetry*, Springer, 1992, pp. 155–245.

[9] D. R. Boyd, O. Eroglu, and M. Kurum, "SCoBi tutorial videos,", 2018, Accessed: 06-Nov-2018.

[10] L. Brocca, L. Ciabatta, C. Massari, S. Camici, and A. Tarpanelli, "Soil moisture for hydrological applications: Open questions and new opportunities," *Water*, vol. 9, no. 2, 2017, p. 140.

[11] M. J. Brodzik, B. Billingsley, T. Haran, B. Raup, and M. H. Savoie, "EASE-Grid 2.0: Incremental but significant improvements for Earth-gridded data sets," *ISPRS International Journal of Geo-Information*, vol. 1, no. 1, 2012, pp. 32–45.

[12] M. Brogioni, M. Caparrini, A. Egido, E. Farres, N. Floury, L. . Guerriero, E. Motte, and N. Pierdicca, "The use of GNSS signals for estimating soil moisture: The Leimon Experiment," *IEEE Proc. of the International Symposium on Geoscience and Remote Sensing-IGARSS*, 2010, pp. 5–7.

[13] A. Camps, H. Park, M. Pablos, G. Foti, C. P. Gommenginger, P.-W. Liu, and J. Judge, "Sensitivity of GNSS-R spaceborne observations to soil moisture and vegetation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 10, 2016, pp. 4730–4742.

[14] A. J. Camps Carmona, *Aplication of Interferometric Radiometry to Earth Observation*, Universitat Politècnica de Catalunya: Barcelona, Spain, 1996.

[15] E. Cardellach, S. Vey, and J. Wickert, "Foreword to the special issue on GNSS reflectometry," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 10, 2016, pp. 4519–4524.

[16] E. Cardellach, J. Wickert, R. Baggen, J. Benito, A. Camps, N. Catarino, B. Chapron, A. Dielacher, F. Fabra, G. Flato, et al., "GNSS transpolar earth reflectometry exploring system (G-TERN): Mission concept," *IEEE access*, vol. 6, 2018, pp. 13980–14018.

[17] H. Carreno-Luengo, A. Camps, J. Querol, and G. Forte, "First results of a GNSS-R experiment from a stratospheric balloon over boreal forests," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 5, 2015, pp. 2652–2663.

[18] H. Carreno-Luengo, A. Camps, P. Via, J. F. Munoz, A. Cortiella, D. Vidal, J. Jané, N. Catarino, M. Hagenfeldt, P. Palomo, et al., "3Cat-2An experimental nanosatellite for GNSS-R earth observation: Mission concept and analysis," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 10, 2016, pp. 4540–4551.

[19] H. Carreno-Luengo, G. Luzi, and M. Crosetto, "Sensitivity of CyGNSS Bistatic Reflectivity and SMAP Microwave Radiometry Brightness Temperature to Geophysical Parameters Over Land Surfaces," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, , no. 99, 2018, pp. 1–16.

[20] T. Castel, A. Beaudoin, N. Floury, T. Le Toan, Y. Caraglio, and J.-F. Barczi, "Deriving forest canopy parameters for backscatter models using the AMAP architectural plant model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 3, 2001, pp. 571–583.

[21] J. Castellví, A. Camps, J. Corbera, and R. Alamús, "3Cat-3/MOTS nanosatellite mission for optical multispectral and GNSS-R earth observation: Concept and analysis," *Sensors*, vol. 18, no. 1, 2018, p. 140.

[22] S. Chan, R. Bindlish, R. Hunt, T. Jackson, and J. Kimball, "Vegetation Water Content,", 2013.

[23] C. Chew, R. Shah, C. Zuffada, G. Hajj, D. Masters, and A. J. Mannucci, "Demonstrating soil moisture remote sensing with observations from the UK TechDemoSat-1 satellite mission," *Geophysical Research Letters*, vol. 43, no. 7, 2016, pp. 3317–3324.

[24] C. Chew and E. Small, "Soil moisture sensing using spaceborne GNSS reflections: Comparison of CYGNSS reflectivity to SMAP soil moisture," *Geophysical Research Letters*, vol. 45, no. 9, 2018, pp. 4049–4057.

[25] C. C. Chew, E. E. Small, K. M. Larson, and V. U. Zavorotny, "Vegetation sensing using GPS-interferometric reflectometry: Theoretical effects of canopy parameters on signal-to-noise ratio data," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, 2015, pp. 2755–2764.

[26] B. Choudhury, T. J. Schmugge, A. Chang, and R. Newton, "Effect of surface roughness on the microwave emission from soils," *Journal of Geophysical Research: Oceans*, vol. 84, no. C9, 1979, pp. 5699–5706.

[27] I. A. Ciampitti, R. W. Elmore, and J. Lauer, "Corn growth and development," *in Kansas State Research and Extension*, 2016.

[28] M. P. Clarizia, N. Pierdicca, F. Costantini, and N. Floury, "Analysis of CYGNSS Data for Soil Moisture Retrieval," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2019.

[29] M. P. Clarizia and C. S. Ruf, "Wind speed retrieval algorithm for the Cyclone Global Navigation Satellite System (CYGNSS) mission," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 8, 2016, pp. 4419–4432.

[30] N. R. Council, *Earth Science and Applications from Space: National Imperatives for The Next Decade And Beyond*, National Academies Press, 2007.

[31] P. de Matthaeis and R. H. Lang, "Microwave scattering models for cylindrical vegetation components," *Progress In Electromagnetics Research*, vol. 55, 2005, pp. 307–333.

[32] R. D. De Roo, Y. Du, F. T. Ulaby, and M. C. Dobson, "A semi-empirical backscattering model at L-band and C-band for a soybean canopy with soil moisture inversion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 4, 2001, pp. 864–872.

[33] R. D. De Roo and F. T. Ulaby, "Bistatic specular scattering from rough dielectric surfaces," *IEEE Transactions on Antennas and Propagation*, vol. 42, no. 2, 1994, pp. 220–231.

[34] M. C. Dobson and F. T. Ulaby, "Active microwave soil moisture research," *IEEE Transactions on Geoscience and Remote Sensing*, , no. 1, 1986, pp. 23–36.

[35] M. C. Dobson, F. T. Ulaby, M. T. Hallikainen, and M. A. El-Rayes, "Microwave dielectric behavior of wet soil-Part II: Dielectric mixing models," *IEEE Transactions on Geoscience and Remote Sensing*, , no. 1, 1985, pp. 35–46.

[36] W. Dorigo, A. Gruber, R. De Jeu, W. Wagner, T. Stacke, A. Loew, C. Albergel, L. Brocca, D. Chung, R. Parinussa, et al., "Evaluation of the ESA CCI soil moisture product using ground-based observations," *Remote Sensing of Environment*, vol. 162, 2015, pp. 380–395.

[37] Y. Du, F. T. Ulaby, and M. C. Dobson, "Sensitivity to soil moisture by active and passive microwave sensors," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 1, 2000, pp. 105–114.

[38] A. Egido, "GNSS reflectometry for land remote sensing applications," *Ph. D. dissertation*, 2013.

[39] A. Egido, M. Caparrini, G. Ruffini, S. Paloscia, E. Santi, L. Guerriero, N. Pierdicca, and N. Floury, "Global navigation satellite systems reflectometry as a remote sensing tool for agriculture," *Remote Sens.*, vol. 4, no. 8, 2012, pp. 2356–2372.

[40] A. Egido, S. Paloscia, E. Motte, L. Guerriero, N. Pierdicca, M. Caparrini, E. Santi, G. Fontanelli, and N. Floury, "Airborne GNSS-R polarimetric measurements for soil moisture and above-ground biomass estimation," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 7, no. 5, 2014, pp. 1522–1532.

[41] D. Entekhabi, G. R. Asrar, A. K. Betts, K. J. Beven, R. L. Bras, C. J. Duffy, T. Dunne, R. D. Koster, D. P. Lettenmaier, D. B. McLaughlin, W. J. Shuttleworth, M. T. V. Genuchten, M. Y. Wei, and E. F. Wood, "An agenda for land surface hydrology research and a call for the Second International Hydrological Decade," *Bull. Am. Meteorol. Soc.*, vol. 80, no. 10, 1999, pp. 2043–2058.

[42] D. Entekhabi, E. G. Njoku, P. E. ONeill, K. H. Kellogg, W. T. Crow, W. N. Edelstein, J. K. Entin, S. D. Goodman, T. J. Jackson, J. Johnson, J. Kimball, J. R. Piepmeier, R. D. Koster, N. Martin, and K. C. McDonald, "The soil moisture active passive (SMAP) mission," *Proc. IEEE*, vol. 98, no. 5, 2010, pp. 704–716.

[43] D. Entekhabi, R. H. Reichle, R. D. Koster, and W. T. Crow, "Performance metrics for soil moisture retrievals and application requirements," *Journal of Hydrometeorology*, vol. 11, no. 3, 2010, pp. 832–840.

[44] D. Entekhabi, I. Rodriguez-Iturbe, and F. Castelli, "Mutual interaction of soil moisture state and atmospheric processes," *Journal of Hydrology*, vol. 184, no. 1-2, 1996, pp. 3–17.

[45] O. Eroglu, D. R. Boyd, and M. Kurum, "Open-sourcing of a SoOp simulator with bistatic vegetation scattering model," *IEEE Proc. of the International Symposium on Geoscience and Remote Sensing-IGARSS*, 2018.

[46] O. Eroglu, D. R. Boyd, and M. Kurum, "SCoBi developers manual,", 2018, Accessed: 06-Nov-2018.

[47] O. Eroglu, D. R. Boyd, and M. Kurum, "SCoBi quick start guide,", 2018, Accessed: 06-Nov-2018.

[48] O. Eroglu, D. R. Boyd, and M. Kurum, "SCoBi users manual,", 2018, Accessed: 06-Nov-2018.

[49] O. Eroglu, D. R. Boyd, and M. Kurum, "SCoBi v1.0.0 release package,", 2018, Accessed: 06-Nov-2018.

[50] O. Eroglu and M. Kurum, "Unveiling CYGNSS Land Signatures for High Spatiotemporal Soil Moisture Estimation,", 2018.

[51] O. Eroglu, M. Kurum, and J. Ball, "Response of GNSS-R on dynamic vegetated terrain conditions," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 5, 2019, pp. 1599–1611.

[52] P. Ferrazzoli, L. Guerriero, N. Pierdicca, and R. Rahmoune, "Forest biomass monitoring with GNSS-R: Theoretical simulations," *Advances in Space Research*, vol. 47, no. 10, 2011, pp. 1823–1832.

[53] L. L. Foldy, "The multiple scattering of waves. I. General theory of isotropic scattering by randomly distributed scatterers," *Physical Review*, vol. 67, no. 3-4, 1945, p. 107.

[54] G. Foti, C. Gommenginger, P. Jales, M. Unwin, A. Shaw, C. Robertson, and J. Rosello, "Spaceborne GNSS reflectometry for ocean winds: First results from the UK TechDemoSat-1 mission," *Geophysical Research Letters*, vol. 42, no. 13, 2015, pp. 5435–5441.

[55] M. Friedl and D. Sulla-Menashe, "MCD12Q1 MODIS/Terra+ Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V006 [Data set]," *NASA EOSDIS Land Processes DAAC. Doi*, vol. 10, 2015.

[56] J. Garrison, B. Nold, Y. Lin, G. Pignotti, J. Piepmeier, M. Vega, M. Fritts, C. DuToit, and J. Knuble, "Recent results on soil moisture remote sensing using P-band signals of opportunity," *2017 International Conference on Electromagnetics in Advanced Applications (ICEAA)*. IEEE, 2017, pp. 1604–1607.

[57] J. L. Garrison, E. Cardellach, S. Gleason, and S. Katzberg, "Foreword to special issue on reflectometry using global navigation satellite systems and other signals of opportunity (GNSS+ R)," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 5, 2014, pp. 1412–1415.

[58] J. L. Garrison, M. Kurum, B. Nold, J. Piepmeier, M. A. Vega, R. Bindlish, and G. Pignotti, "Remote Sensing of Root-Zone Soil Moisture Using I-and P-Band Signals of Opportunity: Instrument Validation Studies," *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2018, pp. 8305–8308.

[59] GCOS, "Global Climate Observing System Essential Climate Variablest,", 2012.

[60] S. Gleason, M. Adjrad, and M. Unwin, "Sensing ocean, ice and land reflected signals from space: results from the UK-DMC GPS reflectometry experiment," *Proceedings of the 2005 ION GNSS Technical Meeting*. Citeseer, 2005, pp. 13–16.

[61] S. Gleason, S. Hodgart, Y. Sun, C. Gommenginger, S. Mackin, M. Adjrad, and M. Unwin, "Detection and processing of bistatically reflected GPS signals from low earth orbit for the purpose of ocean remote sensing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 6, 2005, pp. 1229–1241.

[62] S. Gleason, C. S. Ruf, A. J. OBrien, and D. S. McKague, "The CYGNSS Level 1 calibration algorithm and error analysis based on on-orbit measurements," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 1, 2018, pp. 37–49.

[63] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, 2010, pp. 249–256.

[64] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, "Google Earth Engine: Planetary-scale geospatial analysis for everyone," *Remote Sensing of Environment*, vol. 202, 2017, pp. 18–27.

[65] M. S. Grant, S. T. Acton, and S. J. Katzberg, "Terrain moisture classification using GPS surface-reflected signals," *Remote Sens. IEEE Geosci. Lett.*, vol. 4, no. 1, 2007, pp. 41–45.

[66] L. Guerriero, N. Pierdicca, L. Pulvirenti, and P. Ferrazzoli, "Use of satellite radar bistatic measurements for crop monitoring: A simulation study on corn fields, vol. 5, no. 2," *Remote Sens.*, vol. 5, no. 2, 2013, pp. 864–890.

[67] G. A. Hajj and C. Zuffada, "Theoretical description of a bistatic system for ocean altimetry using the GPS signal," *Radio Science*, vol. 38, no. 5, 2003.

[68] J. Hamaker, J. Bregman, and R. Sault, "Understanding radio polarimetry. I. Mathematical foundations," *Astronomy and Astrophysics Supplement Series*, vol. 117, no. 1, 1996, pp. 137–147.

[69] G. E. Hinton, "Connectionist learning procedures," *Machine learning*, Elsevier, 1990, pp. 555–610.

[70] B. Hofmann-Wellenhof, H. Lichtenegger, and E. Wasle, *GNSS-Global Navigation Satellite Systems: GPS, GLONASS, Galileo, and more*, Springer Science Business Media, 2007.

[71] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural networks*, vol. 2, no. 5, 1989, pp. 359–366.

[72] A. Ishimaru, *Wave propagation and scattering in random media*, vol. 2, Academic press New York, 1978.

[73] T. Jackson, A. Chang, and T. Schmugge, "Active microwave measurements for estimating soil moisture in Oklahoma," 1981.

[74] T. Jackson and T. Schmugge, "Vegetation effects on the microwave emission of soils," *Remote Sensing of Environment*, vol. 36, no. 3, 1991, pp. 203–212.

[75] T. J. Jackson, "III. Measuring surface soil moisture using passive microwave remote sensing," *Hydrological processes*, vol. 7, no. 2, 1993, pp. 139–152.

[76] T. J. Jackson, J. Schmugge, and E. Engman, "Remote sensing applications to hydrology: soil moisture," *Hydrological Sciences Journal*, vol. 41, no. 4, 1996, pp. 517–530.

[77] A. Jarvis, H. I. Reuter, A. Nelson, and E. Guevara, "Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database,", 2008.

[78] A. Joseph, M. Deshpande, P. O'Neill, and L. Miles, "Development of VHF (240–270 MHz) antennas for SoOp (Signal of Opportunity) receiver for 6U cubesat platforms," *2016 Progress in Electromagnetic Research Symposium (PIERS)*. IEEE, 2016, pp. 2530–2531.

[79] M. A. Karam, A. K. Fung, and Y. M. M. Antar, "Electromagnetic wave scattering from some vegetation samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 26, no. 6, 1988, pp. 799–808.

[80] B. Karlik and A. V. Olgac, "Performance analysis of various activation functions in generalized MLP architectures of neural networks," *International Journal of Artificial Intelligence and Expert Systems*, vol. 1, no. 4, 2011, pp. 111–122.

[81] S. J. Katzberg and J. L. Garrison Jr, "Utilizing GPS to determine ionospheric delay over the ocean," 1996.

[82] S. J. Katzberg, O. Torres, M. S. Grant, and D. Masters, "Utilizing calibrated GPS reflected signals to estimate soil reflectivity and dielectric constant: Results from SMEX02," *Remote Sens. Environ.*, vol. 100, no. 1, January 2006, pp. 17–28.

[83] D. E. Kerr, S. Goudsmit, and L. B. Linford, *Propagation of short radio waves*, vol. 24, IET, 1987.

[84] Y. Kerr, P. Waldteufel, P. Richaume, I. Davenport, P. Ferrazzoli, and J. Wigneron, "SMOS level 2 processor soil moisture algorithm theoretical basis document (ATBD)," *SM-ESL (CBSA), CESBIO, Toulouse, SO-TN-ESL-SM-GS-0001, V5. a, 15/03*, 2006.

[85] Y. H. Kerr, P. Waldteufel, J. p. Wigneron, S. Delwart, F. Cabot, J. Boutin, M. j. Escorihuela, J. Font, N. Reul, C. Gruhier, S. E. Juglea, M. R. Drinkwater, A. Hahne, M. Martin-Neira, and S. Mecklenburg, "The SMOS Mission: New tool for monitoring key elements of the global water cycle," *Proc. IEEE*, vol. 98, no. 5, 2010, pp. 666–687.

[86] Y. H. Kerr, P. Waldteufel, J.-P. Wigneron, J. Martinuzzi, J. Font, and M. Berger, "Soil moisture retrieval from space: The Soil Moisture and Ocean Salinity (SMOS) mission," *IEEE Transactions on Geoscience and remote sensing*, vol. 39, no. 8, 2001, pp. 1729–1735.

[87] H. Kim and V. Lakshmi, "Use of Cyclone Global Navigation Satellite System (CYGNSS) observations for estimation of soil moisture," *Geophysical Research Letters*, vol. 45, no. 16, 2018, pp. 8272–8282.

[88] S.-B. Kim, J. J. Van Zyl, J. T. Johnson, M. Moghaddam, L. Tsang, A. Colliander, R. S. Dunbar, T. J. Jackson, S. Jaruwatanadilok, R. West, et al., "Surface soil moisture retrieval using the l-band synthetic aperture radar onboard the soil moisture active–passive satellite and evaluation at core validation sites," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 4, 2017, pp. 1897–1914.

[89] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[90] A. Komjathy, J. Maslanik, V. U. Zavorotny, P. Axelrad, and S. J. Katzberg, "Sea ice remote sensing using surface reflected GPS signals," *IGARSS 2000. IEEE 2000 International Geoscience and Remote Sensing Symposium. Taking the Pulse of the Planet: The Role of Remote Sensing in Managing the Environment. Proceedings (Cat. No. 00CH37120)*. IEEE, 2000, vol. 7, pp. 2855–2857.

[91] M. Kurum, *L-band estimation of forest canopy attenuation by a time-domain analysis of radar backscatter response*, doctoral dissertation, The George Washington University, 2009.

[92] M. Kurum, M. Deshpande, A. T. Joseph, P. E. O'Neill, R. H. Lang, and O. Eroglu, "Development of a coherent bistatic vegetation model for Signal of Opportunity applications at VHF/UHF-bands," *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2017, pp. 4894–4896.

[93] M. Kurum, M. Deshpande, A. T. Joseph, P. E. O'Neill, R. H. Lang, and O. Eroglu, "SCoBi-Veg: A generalized bistatic scattering model of reflectometry from vegetation for Signals of Opportunity applications," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 2, 2018, pp. 1049–1068.

[94] M. Kurum, R. H. Lang, P. E. O'Neill, A. T. Joseph, T. J. Jackson, and M. H. Cosh, "A first-order radiative transfer model for microwave radiometry of forest canopies at L-band," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 9, 2010, pp. 3167–3179.

[95] R. H. Lang, "Electromagnetic backscattering from a sparse distribution of lossy dielectric scatterers," *Radio Science*, vol. 16, no. 01, 1981, pp. 15–30.

[96] R. H. Lang, R. Landry, O. Kavaklioglu, and J.-C. Deguise, "Simulation of microwave backscatter from a red pine stand," *Multispectral and Microwave Sensing of Forestry, Hydrology, and Natural Resources*. International Society for Optics and Photonics, 1995, vol. 2314, pp. 538–548.

[97] K. M. Larson, J. J. Braun, E. E. Small, V. U. Zavorotny, E. D. Gutmann, and A. L. Bilich, "GPS multipath and its relation to near-surface soil moisture content," *Remote Sens. IEEE J. Sel. Top. Appl. Earth Obs.*, vol. 3, no. 1, 2010, pp. 91–99.

[98] K. M. Larson, E. D. Gutmann, V. U. Zavorotny, J. J. Braun, M. W. Williams, and F. G. Nievinski, "Can we measure snow depth with GPS receivers?," *Geophysical Research Letters*, vol. 36, no. 17, 2009.

[99] K. M. Larson and E. E. Small, "Using GPS to study the terrestrial water cycle," *Eos, Transactions American Geophysical Union*, vol. 94, no. 52, 2013, pp. 505–506.

[100] K. M. Larson, E. E. Small, J. J. Braun, and V. U. Zavorotny, "Environmental sensing," *A revolution in GNSS applications. Inside GNSS*, 2014, pp. 36–46.

[101] K. M. Larson, E. E. Small, E. D. Gutmann, A. L. Bilich, J. J. Braun, and V. U. Zavorotny, "Use of GPS receivers as a soil moisture network for water cycle studies," *Geophysical Research Letters*, vol. 35, no. 24, 2008.

[102] M. Lax, "Multiple scattering of waves," *Reviews of Modern Physics*, vol. 23, no. 4, 1951, p. 287.

[103] J. Leese, T. Jackson, A. Pitman, and P. Dirmeyer, "GEWEX/BAHC international workshop on soil moisture monitoring, analysis, and prediction for hydrometerological and hydroclimatological applications," *Bull. Am. Am. Meteorol. Soc.*, vol. 82, no. 7, 2001, pp. 1423–1430.

[104] D. M. LeVine, R. Meneghini, R. H. Lang, and S. S. Seker, "Scattering from arbitrarily oriented dielectric disks in the physical optics regime," *J. Opt. Soc. Am.*, vol. 73, no. 10, 1983, pp. 1255–1262.

[105] D. M. LeVine, A. Schneider, R. H. Lang, and H. G. Carter, "Scattering from thin dielctric disks," *IEEE Trans. Antennas Propag.*, vol. 33, no. 12, 1985, pp. 1410–1413.

[106] P. Liang, L. E. Pierce, and M. Moghaddam, "Radiative transfer model for microwave bistatic scattering from forest canopies," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 11, 2005, pp. 2470–2483.

[107] B. Lin, S. J. Katzberg, J. L. Garrison, and B. A. Wielicki, "Relationship between GPS signals reflected from sea surfaces and surface winds: Modeling results and comparisons with aircraft measurements," *Journal of Geophysical Research: Oceans*, vol. 104, no. C9, 1999, pp. 20713–20727.

[108] Y.-C. Lin and K. Sarabandi, "A Monte Carlo coherent scattering model for forest canopies using fractal-generated trees," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 1, 1999, pp. 440–451.

[109] S. T. Lowe, J. L. LaBrecque, C. Zuffada, L. J. Romans, L. E. Young, and G. A. Hajj, "First spaceborne observation of an Earth-reflected GPS signal," *Radio Science*, vol. 37, no. 1, 2002, pp. 1–28.

[110] A. Ludwig, "The definition of cross polarization," *IEEE Transactions on Antennas and Propagation*, vol. 21, no. 1, 1973, pp. 116–119.

[111] M. Martin-Neira et al., "A passive reflectometry and interferometry system (PARIS): Application to ocean altimetry," *ESA journal*, vol. 17, no. 4, 1993, pp. 331–355.

[112] D. Masters, *Surface remote sensing applications of GNSS bistatic radar: Soil moisture and aircraft altimetry*, Citeseer, 2004.

[113] D. Masters, P. Axelrad, and S. Katzberg, "Initial results of land-reflected GPS bistatic radar measurements in SMEX02," *Remote sensing of environment*, vol. 92, no. 4, 2004, pp. 507–520.

[114] D. Masters, V. Zavorotny, S. Katzberg, and W. Emery, "GPS signal scattering from land for moisture content determination," *Geoscience and Remote Sensing Symposium, 2000. Proceedings. IGARSS 2000. IEEE 2000 International*, vol. 7, pp., 2000, pp. 3090–3092.

[115] V. L. Mironov, R. D. De Roo, and I. V. Savin, "Temperature-dependable microwave dielectric model for an Arctic soil," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 6, 2010, pp. 2544–2556.

[116] H. Mott, "Polarization in antennas and radar," *New York, Wiley-Interscience, 1986. 312 p.*, 1986.

[117] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010, pp. 807–814.

[118] S. V. Nghiem, C. Zuffada, R. Shah, C. Chew, S. T. Lowe, A. J. Mannucci, E. Cardellach, G. R. Brakenridge, G. Geller, and A. Rosenqvist, "Wetland monitoring with Global Navigation Satellite System reflectometry," *Earth Sp. Sci.*, vol. 4, no. 1, 2017, pp. 16–39.

[119] F. G. Nievinski and K. M. Larson, "Forward modeling of GPS multipath for near-surface reflectometry and positioning applications," *GPS solutions*, vol. 18, no. 2, 2014, pp. 309–322.

[120] F. G. Nievinski and K. M. Larson, "An open source GPS multipath simulator in Matlab/Octave," *Gps Solutions*, vol. 18, no. 3, 2014, pp. 473–481.

[121] E. G. Njoku and D. Entekhabi, "Passive microwave remote sensing of soil moisture," *Journal of hydrology*, vol. 184, no. 1-2, 1996, pp. 101–129.

[122] E. G. Njoku and P. E. O'Neill, "Multifrequency microwave radiometer measurements of soil moisture," *IEEE Transactions on Geoscience and Remote Sensing*, , no. 4, 1982, pp. 468–475.

[123] P. O'Neill, M. Kurum, A. Joseph, J. Fuchs, P. Young, M. Cosh, and R. Lang, "L-band active / passive time series measurements over a growing season using the ComRAD ground-based SMAP simulator," *IEEE Proc. of the International Symposium on Geoscience and Remote Sensing-IGARSS*, 2013, pp. 37–40.

[124] P. ONeill, S. Chan, E. Njoku, T. Jackson, and R. Bindlish, "Soil moisture active passive (SMAP) algorithm theoretical basis document Level 2 & 3 soil moisture (passive) data products," *Jet Propulsion Laboratory. California Institute of Technology*, 2014.

[125] S. Paloscia, P. Pampaloni, S. Pettinato, P. Poggi, and E. Santi, "The retrieval of soil moisture from ENVISAT/ASAR data," *EARSeL eProceedings*, vol. 4, no. 1, 2005, pp. 44–51.

[126] S. Paloscia, S. Pettinato, E. Santi, C. Notarnicola, L. Pasolli, and A. Reppucci, "Soil moisture mapping using Sentinel-1 images: Algorithm and preliminary validation," *Remote Sensing of Environment*, vol. 134, 2013, pp. 234–248.

[127] C. Pathe, W. Wagner, D. Sabel, M. Doubkova, and J. B. Basara, "Using ENVISAT ASAR global mode data for surface soil moisture retrieval over Oklahoma, USA," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 2, 2009, pp. 468–480.

[128] J.-F. Pekel, A. Cottam, N. Gorelick, and A. S. Belward, "High-resolution mapping of global surface water and its long-term changes," *Nature*, vol. 540, no. 7633, 2016, p. 418.

[129] R. R. Picard and R. D. Cook, "Cross-validation of regression models," *Journal of the American Statistical Association*, vol. 79, no. 387, 1984, pp. 575–583.

[130] N. Pierdicca, L. Guerriero, R. Giusto, M. Brogioni, and A. Egido, "SAVERS: A simulator of GNSS reflections from bare and vegetated soils," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 10, 2014, pp. 6542–6554.

[131] F. Pukelsheim, "The three sigma rule," *The American Statistician*, vol. 48, no. 2, 1994, pp. 88–91.

[132] N. Rodriguez-Alvarez, X. Bosch-Lluis, A. Camps, A. Aguasca, M. Vall-Llossera, E. Valencia, I. Ramos-Perez, and H. Park, "Review of crop growth and soil moisture monitoring from a ground-based instrument implementing the Interference Pattern GNSS-R Technique," *Radio Sci.*, vol. 46, no. 6, 2011.

[133] N. Rodriguez-Alvarez, X. Bosch-Lluis, A. Camps, M. Vall-Llossera, E. Valencia, J. F. Marchan-Hernandez, and I. Ramos-Perez, "Soil moisture retrieval using GNSS-R techniques: Experimental results over a bare soil field," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, 2009, pp. 3616–3624.

[134] N. Rodriguez-Alvarez, A. Camps, M. Vall-Llossera, X. Bosch-Lluis, A. Monerris, I. Ramos-Perez, E. Valencia, J. Martinez-Fernandez, G. Baroncini-Turricchia, C. Perez-Gutierrez, N. Sanchez, and J. F. Marchan-Hernandez, "Land geophysical parameters retrieval using the interference pattern GNSS-R technique," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 1, 2011, pp. 71–84.

[135] N. Rodriguez-Alvarez, E. Podest, K. Jensen, and K. C. McDonald, "Classifying Inundation in a Tropical Wetlands Complex with GNSS-R," *Remote Sensing*, vol. 11, no. 9, 2019, p. 1053.

[136] R. Rose, S. Gleason, and C. Ruf, "NASA CYGNSS Mission Update; A Pathfinder for Operational GNSS Scatterometry Remote Sensing Applications," 2017.

[137] A. Rosenqvist, M. Shimada, N. Ito, and M. Watanabe, "ALOS/PALSAR: Pathfinder mission for global-scale monitoring of the environment," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 11, 2007, pp. 3307–3316.

[138] C. Ruf, P. Chang, M. Clarizia, S. Gleason, Z. Jelenak, J. Murray, M. Morris, S. Musko, D. Posselt, D. Provost, et al., "Cygnss handbook," *Ann Arbor, MI, Michigan Pub*, vol. 154, 2016.

[139] C. Ruf, J. Redfern, T. Butler, D. McKague, and S. Gleason, "Level 1B DDM Calibration Algorithm Theoretical Basis Document Rev. 2,", 2018, Last accessed 24 July 2019.

[140] C. Ruf, M. Unwin, J. Dickinson, R. Rose, D. Rose, M. Vincent, and A. Lyons, "CYGNSS: Enabling the future of hurricane prediction [remote sensing satellites]," *IEEE Geoscience and Remote Sensing Magazine*, vol. 1, no. 2, 2013, pp. 52–67.

[141] C. S. Ruf, R. Atlas, P. S. Chang, M. P. Clarizia, J. L. Garrison, S. Gleason, S. J. Katzberg, Z. Jelenak, J. T. Johnson, S. J. Majumdar, and A. Obrien, "New ocean winds satellite mission to probe hurricanes and tropical convection," *Bulletin of the American Meteorological Society*, vol. 97, no. 3, 2016, pp. 385–395.

[142] C. S. Ruf, C. Chew, T. Lang, M. G. Morris, K. Nave, A. Ridley, and R. Balasubramaniam, "A new paradigm in earth environmental monitoring with the CYGNSS small satellite constellation," *Scientific reports*, vol. 8, no. 1, 2018, p. 8782.

[143] C. S. Ruf, S. Gleason, Z. Jelenak, S. Katzberg, A. Ridley, R. Rose, J. Scherrer, and V. Zavorotny, "The CYGNSS nanosatellite constellation hurricane mission," *2012 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2012, pp. 214–216.

[144] C. S. Ruf, S. Gleason, and D. S. McKague, "Assessment of CYGNSS wind speed retrieval uncertainty," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 1, 2018, pp. 87–97.

[145] E. Santi, S. Paloscia, P. Pampaloni, S. Pettinato, and P. Poggi, "Retrieval of soil moisture from ENVISAT ASAR images: a comparison of inversion algorithms," *Proceedings of the 2004 Envisat & ERS Symposium (ESA SP-572)*, 2004, pp. 6–10.

[146] G. L. Schaefer, M. H. Cosh, and T. J. Jackson, "The USDA natural resources conservation service soil climate analysis network (SCAN)," *Journal of Atmospheric and Oceanic Technology*, vol. 24, no. 12, 2007, pp. 2073–2077.

[147] S. S. Seker and A. Schneider, "Electromagnetic scattering from a dielectric cylinder of finite length," *IEEE Trans. Antennas Propag.*, vol. 36, no. 2, 1988, pp. 303–307.

[148] R. Shah, J. L. Garrison, A. Egido, and G. Ruffini, "Bistatic radar measurements of significant wave height using signals of opportunity in L-, S-, and Ku-bands," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 2, 2015, pp. 826–841.

[149] R. Shah, J. L. Garrison, and M. S. Grant, "Demonstration of bistatic radar for ocean remote sensing using communication satellite signals," *IEEE Geoscience and Remote Sensing Letters*, vol. 9, no. 4, 2011, pp. 619–623.

[150] R. Shah, X. Xu, S. Yueh, C. S. Chae, K. Elder, B. Starr, and Y. Kim, "Remote sensing of snow water equivalent using P-band coherent reflection," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 3, 2017, pp. 309–313.

[151] R. Shah, C. Zuffada, C. Chew, M. Lavalle, X. Xu, and A. Azemati, "Modeling bistatic scattering signatures from sources of opportunity in P-Ka bands," *2017 International Conference on Electromagnetics in Advanced Applications (ICEAA)*. IEEE, 2017, pp. 1684–1687.

[152] E. E. Small, K. M. Larson, C. C. Chew, J. Dong, and T. E. Ochsner, "Validation of GPS-IR soil moisture retrievals: Comparison of different algorithms to remove vegetation effects," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 10, 2016, pp. 4759–4770.

[153] E. E. Small, K. M. Larson, and W. K. Smith, "Normalized microwave reflection index: validation of vegetation water content estimates from Montana grasslands," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 5, 2014, pp. 1512–1521.

[154] A. Smith, J. P. Walker, A. W. Western, R. Young, K. Ellett, R. Pipunic, R. Grayson, L. Siriwardena, F. Chiew, and H. Richter, "The Murrumbidgee soil moisture monitoring network data set," *Water Resources Research*, vol. 48, no. 7, 2012.

[155] L. Thirion-Lefevre, E. Colin-Koeniguer, and C. Dahon, "Bistatic scattering from forest components. Part I: coherent polarimetric modelling and analysis of simulated results," *Waves in Random and Complex Media*, vol. 20, no. 1, February 2010, pp. 36–61.

[156] L. Tsang, J. A. Kong, and R. T. Shin, "Theory of microwave remote sensing," 1985.

[157] S. Twomey, *Introduction to the mathematics of inversion in remote sensing and indirect measurements*, vol. 3, Elsevier, 2013.

[158] F. T. Ulaby and P. P. Batlivala, "Optimum radar parameters for mapping soil moisture," *IEEE Transactions on Geoscience Electronics*, vol. 14, no. 2, 1976, pp. 81–93.

[159] F. T. Ulaby, P. P. Batlivala, and M. C. Dobson, "Microwave backscatter dependence on surface roughness, soil moisture, and soil texture: Part I-bare soil," *IEEE Transactions on Geoscience Electronics*, vol. 16, no. 4, 1978, pp. 286–295.

[160] F. T. Ulaby, D. G. Long, W. J. Blackwell, C. Elachi, A. K. Fung, C. Ruf, K. Sarabandi, H. A. Zebker, and J. Van Zyl, *Microwave radar and radiometric remote sensing*, vol. 4, University of Michigan Press Ann Arbor, 2014.

[161] F. T. Ulaby, R. K. Moore, and A. K. Fung, "Microwave remote sensing active and passive," 2015.

[162] M. Unwin, S. Gleason, and M. Brennan, "The space GPS reflectometry experiment on the UK disaster monitoring constellation satellite," *Proceedings of ION-GPS/GNSS*. Portland, 2003.

[163] M. Unwin, P. Jales, J. Tye, C. Gommenginger, G. Foti, and J. Rosello, "Spaceborne GNSS-reflectometry on TechDemoSat-1: Early mission operations and exploitation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 10, 2016, pp. 4525–4539.

[164] C. Utku and R. H. Lang, "Coherence effects in L-Band active and passive remote sensing of quasi-periodic corn canopies," *General Assembly and Scientific Symposium, 2011 XXXth URSI*, 2011, pp. 1–4.

[165] E. Vermote, C. Justice, I. Csiszar, J. Eidenshink, R. Myneni, F. Baret, E. Masuoka, R. Wolfe, and M. Claverie, "NOAA Climate Data Record (CDR) of normalized Difference Vegetation Index (NDVI), Version 4," *NOAA Natl. Clim. Data Cent*, 2014.

[166] E. Vermote and R. Wolfe, "MOD09GA MODIS/Terra Surface Reflectance Daily L2G Global 1 km and 500 m SIN Grid V006," *NASA EOSDIS Land Processes DAAC. Available online: https://lpdaac. usgs. gov/dataset_ discovery/modis/modis_products_table/mod09ga_v006 (accessed on 16 October 2016)*, 2015.

[167] A. G. Voronovich and V. U. Zavorotny, "Bistatic radar equation for signals of opportunity revisited," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 4, 2017, pp. 1959–1968.

[168] W. Wagner, G. Blschl, P. Pampaloni, J. c. Calvet, B. Bizzarri, J. p. Wigneron, and Y. Kerr, "Operational readiness of microwave remote sensing of soil moisture for hydrologic applications," *Hydrol. Res.*, vol. 38, no. 1, 2007, pp. 1–20.

[169] W. Wagner, S. Hahn, R. Kidd, T. Melzer, Z. Bartalis, S. Hasenauer, J. Figa, P. de Rosnay, A. Jann, S. Schneider, et al., "The ASCAT Soil Moisture Product: Specifications, Validation, Results, and Emerging Applications, submitted to," *Meteorologische Zeitschrift*, 2012.

[170] J. R. Wang and T. J. Schmugge, "An empirical model for the complex dielectric permittivity of soils as a function of water content," *IEEE Transactions on Geoscience and Remote Sensing*, , no. 4, 1980, pp. 288–295.

[171] W. Wasylkiwskyj, "Response of an antenna to arbitrary incident fields," *2005 IEEE Antennas and Propagation Society International Symposium*. IEEE, 2005, vol. 3, pp. 39–42.

[172] J. Wen, Q. Liu, Q. Xiao, Q. Liu, D. You, D. Hao, S. Wu, and X. Lin, "Characterizing land surface anisotropic reflectance over rugged terrain: a review of concepts and recent developments," *Remote Sensing*, vol. 10, no. 3, 2018, p. 370.

[173] J.-P. Wigneron, T. Schmugge, A. Chanzy, J.-C. Calvet, and Y. Kerr, "Use of passive microwave remote sensing to monitor soil moisture," 1998.

[174] X. Wu and S. Jin, "GNSS-Reflectometry: Forest canopies polarization scattering properties and modeling," *Adv. Sp. Res.*, vol. 54, no. 5, 2014, pp. 863–870.

[175] X. Wu, S. Jin, and J. Xia, "A forward GPS multipath simulator based on the vegetation radiative transfer equation model," *Sensors*, vol. 17, no. 6, 2017, p. 1291.

[176] Y. Xie, Z. Sha, M. Yu, Y. Bai, and L. Zhang, "A comparison of two models with Landsat data for estimating above ground grassland biomass in Inner Mongolia, China," *Ecological Modelling*, vol. 220, no. 15, 2009, pp. 1810–1818.

[177] S. Yueh, R. Shah, X. Xu, K. Elder, C. S. Chae, S. Margulis, G. Liston, M. Durand, and C. Derksen, "HydroCube mission concept: P-Band signals of opportunity for remote sensing of snow and root zone soil moisture," *Sensors, Systems, and Next-Generation Satellites XXI*. International Society for Optics and Photonics, 2017, vol. 10423, p. 104230L.

[178] V. U. Zavorotny, S. Gleason, E. Cardellach, and A. Camps, "Tutorial on remote sensing using GNSS bistatic radar of opportunity," *IEEE Geoscience and Remote Sensing Magazine*, vol. 2, no. 4, 2014, pp. 8–45.

[179] V. U. Zavorotny and A. G. Voronovich, "Scattering of GPS signals from the ocean with wind remote sensing application," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 2, 2000, pp. 951–964.

[180] C. Zuffada, C. Chew, and S. Nghiem, "GNSS-R algorithms for wetlands observations," *Proceeding of IEEE IGARSS*, 2017.

APPENDIX A

NORMALIZED ”VOLTAGE” PATTERN

In this appendix, the normalized voltage pattern for both linear and circular polarization basis will be expressed. Due to reciprocity, the following arguments are equally applicable for antennas in transmit and receive modes.

In practice, antennas cannot be constructed to produce pure polarization states [116]. It is important to decompose the radiated field into two orthonormal polarization states. The orthogonal polarization states are commonly chosen along the linear or circular basis vectors. For linear basis, co-polarized and cross-polarized patterns are defined for each port as

$$\mathbf{g}_X = g_{XX}\hat{\mathbf{u}}_1^X + g_{XY}\hat{\mathbf{u}}_2^Y \tag{A.1a}$$

$$\mathbf{g}_Y = g_{YX}\hat{\mathbf{u}}_1^X + g_{YY}\hat{\mathbf{u}}_2^Y \tag{A.1b}$$

where g_{XX} is the co-polarized voltage pattern for port 1 while g_{XY} is the cross-polarized voltage pattern for the same port. The quantity g_{YY} is the co-polarized voltage pattern for port 2 while g_{YX} is the cross-polarized voltage pattern for the same port. The voltage patterns are complex. The cross-polarized patterns represent the crosstalk between antenna ports. The unit vectors $\hat{\mathbf{u}}_1^X$ and $\hat{\mathbf{u}}_2^Y$ are also complex and are associated with co-polarization and cross-polarization states.

In linear basis, the polarization vectors are chosen according to Ludwig's third definition of polarization to indicate the directions of co-polarization and cross-polarization [110]. They are rotated versions of spherical vectors defined in antenna coordinates and can be written as

$$\hat{\mathbf{u}}_1^X = cos\varphi\hat{\theta} - sin\varphi\hat{\varphi} \tag{A.2a}$$

$$\hat{\mathbf{u}}_2^Y = sin\varphi\hat{\theta} + cos\varphi\hat{\varphi} \tag{A.2b}$$

where $\hat{\theta}$ and $\hat{\varphi}$ are the unit vectors in spherical coordinates of the antenna. Thus, the normalized voltage pattern matrix for the linearly polarized antenna is given as

$$\underline{\underline{g}} = \begin{bmatrix} g_{XX} & g_{XY} \\ g_{YX} & g_{YY} \end{bmatrix} \tag{A.3}$$

For RHCP as the reference polarization and LHCP as the cross-polarization, the unit polarization vectors are

$$\hat{\mathbf{u}}_1^R = \frac{1}{\sqrt{2}}(\hat{\mathbf{u}}_1^X - i\hat{\mathbf{u}}_2^Y) \tag{A.4a}$$

$$\hat{\mathbf{u}}_2^L = \frac{1}{\sqrt{2}}(\hat{\mathbf{u}}_1^X + i\hat{\mathbf{u}}_2^Y) \tag{A.4b}$$

Similarly, for circular polarization basis, co-polarized and cross-polarized patterns can be written for each port as

$$g_R = g_{RR}\hat{\mathbf{u}}_1^R + g_{RL}\hat{\mathbf{u}}_2^L \tag{A.5a}$$

$$g_L = g_{LR}\hat{\mathbf{u}}_1^R + g_{LL}\hat{\mathbf{u}}_2^L \tag{A.5b}$$

where the subscripts/superscripts R and L denotes RHCP and LHCP, respectively. The normalized voltage pattern matrix for the circularly polarized antenna is given as

$$\underline{\underline{g}} = \begin{bmatrix} g_{RR} & g_{RL} \\ g_{LR} & g_{LL} \end{bmatrix} \tag{A.6}$$

where the diagonal elements are co-polarized patterns while the off-diagonal elements are cross-polarized patterns. The elements are all complex.

APPENDIX B

POLARIZATION BASIS TRANSFORMATION

An arbitrary-polarized electric field can always be expressed by a linear combination of two orthogonal (basis) fields that are usually defined in one coordinate system [8]. Since bistatic scattering involves radiation, scattering, and reception in various coordinate systems, it is necessary to transform polarization basis vectors in one coordinate system to another one. In this appendix, we will consider two configurations: transmitreceive and transmitscatteringreceive as depicted in Figure B.1.

A. TransmitReceive Configuration

Let us first consider two arbitrary-oriented antenna coordinate systems that are denoted by $T : (x_t, y_t, z_t)$ and $R : (x_r, y_r, z_r)$ as shown in Figure B.1(a). The unit vectors $\hat{\mathbf{u}}_{t1}$ and $\hat{\mathbf{u}}_{t2}$ are complex orthonormal and defined in the transmit antenna system T while the unit vectors $\hat{\mathbf{u}}_{r1}$ and $\hat{\mathbf{u}}_{r2}$ are complex orthonormal and defined in the receive antenna system R. An arbitrary electrical field can be decomposed into orthogonal components in both systems as

$$\boldsymbol{E} = E_{t1}\hat{\mathbf{u}}_{t1} + E_{t2}\hat{\mathbf{u}}_{t2} \tag{B.1a}$$

$$\boldsymbol{E} = E_{r1}\hat{\mathbf{u}}_{r1} + E_{r2}\hat{\mathbf{u}}_{r2} \tag{B.1b}$$

The components of the field in both coordinates are related as

$$E_{r1} = \boldsymbol{E} \cdot \hat{\mathbf{u}}^*_{r1} = E_{t1}\left(\hat{\mathbf{u}}_{t1} \cdot \hat{\mathbf{u}}^*_{r1}\right) + E_{t2}\left(\hat{\mathbf{u}}_{t2} \cdot \hat{\mathbf{u}}^*_{r1}\right) \tag{B.2a}$$

$$E_{r2} = \boldsymbol{E} \cdot \hat{\mathbf{u}}^*_{r2} = E_{t1}\left(\hat{\mathbf{u}}_{t1} \cdot \hat{\mathbf{u}}^*_{r2}\right) + E_{t2}\left(\hat{\mathbf{u}}_{t2} \cdot \hat{\mathbf{u}}^*_{r2}\right) \tag{B.2b}$$

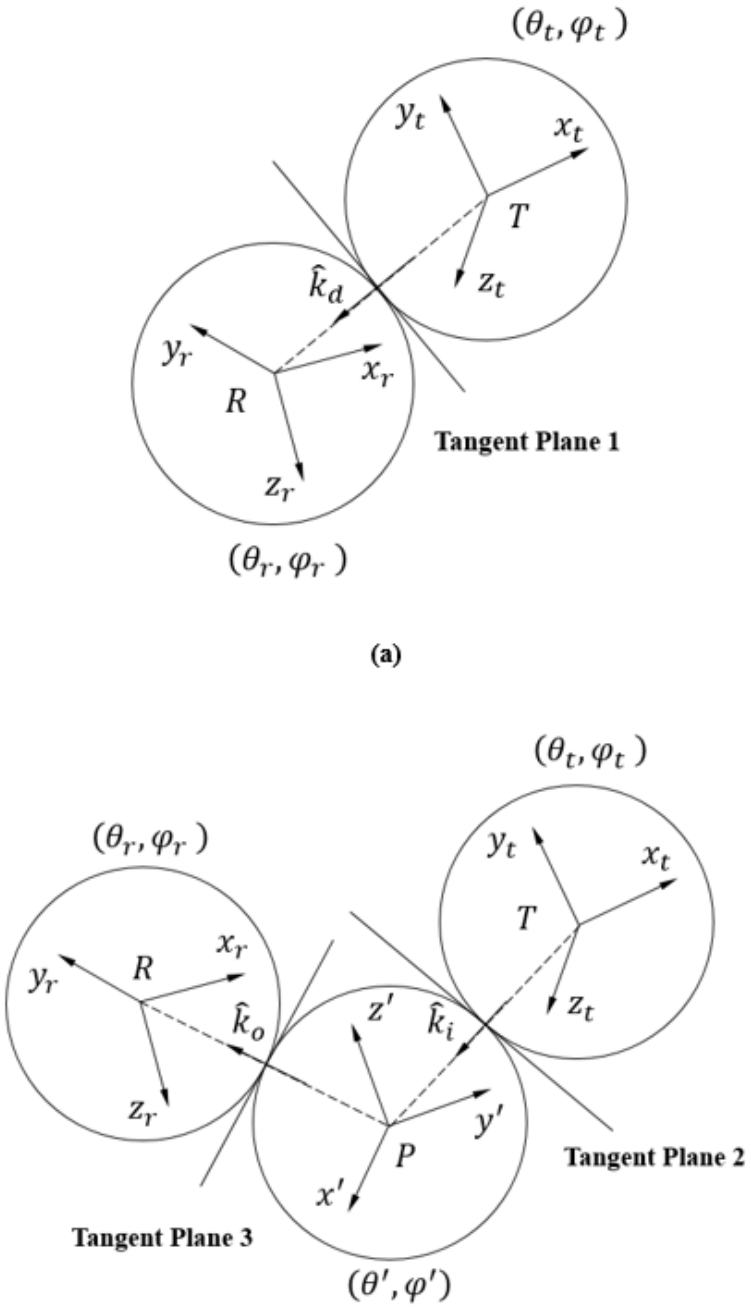

Figure B.1: Change of polarization basis (a) between transmitter and receiver and (b) between transmitter, ground, and receiver.

The above equations can be cast into a matrix form as

$$\begin{bmatrix} E_{r1} \\ \\ E_{r2} \end{bmatrix} = \underline{\underline{u}}_{t\to r} \begin{bmatrix} E_{t1} \\ \\ E_{t2} \end{bmatrix} \tag{B.3a}$$

where the transformation matrix $\underline{\underline{u}}_{t\to r}$ is given by

$$\underline{\underline{u}}_{t\to r} = \begin{bmatrix} \hat{\mathbf{u}}_{t1} \cdot \hat{\mathbf{u}}^*_{r1} & \hat{\mathbf{u}}_{t2} \cdot \hat{\mathbf{u}}^*_{r1} \\ \\ \hat{\mathbf{u}}_{t1} \cdot \hat{\mathbf{u}}^*_{r2} & \hat{\mathbf{u}}_{t2} \cdot \hat{\mathbf{u}}^*_{r2} \end{bmatrix} \tag{B.3b}$$

where it transforms polarization components of the system T to those in the system R. The matrix is, in fact, a unitary matrix that conserves the total power of a wave, that is, the norm of the electric field vector remains invariant under change of basis [8]. As a result, the inverse of the transformation matrix (or transformation from R to T) is

$$\underline{\underline{u}}_{r\to t} = \underline{\underline{u}}^{-1}_{t\to r} = \underline{\underline{u}}^{H}_{t\to r} \tag{B.4}$$

where superscript 1 and H are the inverse and the Hermitian (or conjugate) transpose of a matrix, respectively.

Now let us consider the unit spherical vectors (orthogonal sets) in both coordinate systems. The direction of propagation is denoted by $\hat{\mathbf{k}}_d$ vector that connects origins of both coordinate systems. It is evident that the transmit antenna spherical vectors ($\hat{\boldsymbol{\theta}}_t$, $\hat{\boldsymbol{\varphi}}_t$) lie in the same tangential plane (tangent plane 1), where the receive antenna spherical vectors ($\hat{\boldsymbol{\theta}}_r$, $\hat{\boldsymbol{\varphi}}_r$) lie, and both of which are normal to $\hat{\mathbf{k}}_d$, as illustrated in Figure B.1(a), due to the following orthogonality relationship:

$$\hat{\boldsymbol{\varphi}}_t \times \hat{\mathbf{k}}_d = \hat{\boldsymbol{\theta}}_t \quad \text{and } \hat{\mathbf{k}}_d \times \hat{\boldsymbol{\theta}}_t = \hat{\boldsymbol{\varphi}}_t \tag{B.5a}$$

$$\hat{\mathbf{k}}_d \times \hat{\boldsymbol{\varphi}}_r = \hat{\boldsymbol{\theta}}_r \quad \text{and } \hat{\boldsymbol{\theta}}_r \times \hat{\mathbf{k}}_d = \hat{\boldsymbol{\varphi}}_r \tag{B.5b}$$

The linear and circular basis vectors (defined in (A.2) and (A.4)] also lie on the same tangent plane since they are rotated versions of spherical vectors defined in their respective coordinate systems. As a result, the transformation matrix, $\underline{\underline{u}}_{t\to r}$, is, in fact, a rotation matrix between any polarization bases (within the system T or R and between the systems T and R), defined in the same tangential plane. In other words, any polarization states defined along a direct path between the origins of two arbitrary coordinate systems are related via a rotation matrix. One can go from one polarization basis to another one by simply substituting the unit complex polarization vector pairs [i.e., $(\hat{\boldsymbol{\theta}}_t, \hat{\boldsymbol{\varphi}}_t), (\hat{\mathbf{u}}_{t1}^R, \hat{\mathbf{u}}_{t2}^L), (\hat{\mathbf{u}}_{t1}^X, \hat{\mathbf{u}}_{t2}^Y), (\hat{\boldsymbol{\theta}}_r, \hat{\boldsymbol{\varphi}}_r), (\hat{\mathbf{u}}_{r1}^R, \hat{\mathbf{u}}_{r2}^L)$, and $(\hat{\mathbf{u}}_{r1}^X, \hat{\mathbf{u}}_{r2}^Y)$] in the rotation matrix given in (B.3b).

B. TransmitScatteringReceive Configuration

Now, we add an intermediate coordinate system (x', y', z') to represent local scattering processes between transmit and receive antennas, as shown in Figure B.1(b). In this configuration, the transformation involves two rotation matrices; one is from the transmit antenna to the ground system and the other one is from the ground system to the receive antenna. The incoming and outgoing propagation vectors $\hat{\mathbf{k}}_i$ and $\hat{\mathbf{k}}_o$ can be defined and be written in local coordinates by

$$-\hat{\mathbf{k}}_i = \hat{\mathbf{x}}^{'} \sin\theta_i^{'} \cos\varphi_i^{'} + \hat{\mathbf{y}}^{'} \sin\theta_i^{'} \sin\varphi_i^{'} + \hat{\mathbf{z}}^{'} \cos\theta_i^{'} \tag{B.6a}$$

$$\hat{\mathbf{k}}_o = \hat{\mathbf{x}}^{'} \sin\theta_o^{'} \cos\varphi_o^{'} + \hat{\mathbf{y}}^{'} \sin\theta_o^{'} \sin\varphi_o^{'} + \hat{\mathbf{z}}^{'} \cos\theta_o^{'} \tag{B.6b}$$

where the angles $(\theta_i^{'}, \varphi_i^{'})$ and $(\theta_o^{'}, \varphi_o^{'})$ represent the angle of incidence and the angle of scattering in the local coordinate system, respectively. The propagation vector, $\hat{\mathbf{k}}_i$, is normal to both the transmit antenna spherical vectors $(\hat{\boldsymbol{\theta}}_t, \hat{\boldsymbol{\varphi}}_t)$ and the local spherical vectors $(\hat{\boldsymbol{\theta}}_i^{'}, \hat{\boldsymbol{\varphi}}_i^{'})$ along the angle $(\theta_i^{'}, \varphi_i^{'})$ while the propagation vector, $\hat{\mathbf{k}}_o$, is normal to both the receive antenna spherical vectors $(\hat{\boldsymbol{\theta}}_r, \hat{\boldsymbol{\varphi}}_r)$ and the local spherical vectors $(\hat{\boldsymbol{\theta}}_o^{'}, \hat{\boldsymbol{\varphi}}_o^{'})$ along the angle $(\theta_o^{'}, \varphi_o^{'})$. Due to these orthogonality relationships, we can define a tangent plane (tangent plane 2) between the transmit antenna and the ground and can define another tangent plane (tangent plane 3) between the receive antenna and the ground. The unit vectors $(\hat{\mathbf{u}}_{t1}, \hat{\mathbf{u}}_{t2})$ and $(\hat{\mathbf{u}}_{r1}, \hat{\mathbf{u}}_{r2})$ also lie in the tangent planes 2 and 3, respectively. Thus, the rotation matrices from the transmit antenna to the ground and the ground to the receive antenna can be written as

$$\underline{\underline{u}}_{t\to g}\left(\hat{\mathbf{k}}_i\right) = \begin{bmatrix} \hat{\mathbf{u}}_{t1} \cdot \hat{\mathbf{v}}_i^* & \hat{\mathbf{u}}_{t2} \cdot \hat{\mathbf{v}}_i^* \\ \hat{\mathbf{u}}_{t1} \cdot \hat{\mathbf{h}}_i^* & \hat{\mathbf{u}}_{t2} \cdot \hat{\mathbf{h}}_i^* \end{bmatrix} \tag{B.7a}$$

and

$$\underline{\underline{u}}_{g\to r}\left(\hat{\mathbf{k}}_O\right) = \begin{bmatrix} \hat{\mathbf{v}}_o \cdot \hat{\mathbf{u}}_{r1}^* & \hat{\mathbf{h}}_o \cdot \hat{\mathbf{u}}_{r1}^* \\ \hat{\mathbf{v}}_o \cdot \hat{\mathbf{u}}_{r2}^* & \hat{\mathbf{h}}_o \cdot \hat{\mathbf{u}}_{r2}^* \end{bmatrix} \tag{B.7b}$$

where $(\hat{\mathbf{u}}_{t1}, \hat{\mathbf{u}}_{t2})$ and $(\hat{\mathbf{u}}_{r1}, \hat{\mathbf{u}}_{r2})$ can be the linearly polarized or circularly polarized unit vectors as defined in (A.2) and (A.4). The local horizontal polarization vector, $\hat{\mathbf{h}}_o$, for scattered wave is taken parallel to the local ground $(x'y')$ plane. More specifically

$$\hat{\mathbf{h}}_o = \frac{\hat{\mathbf{n}} \times \hat{\mathbf{k}}_o}{\left|\hat{\mathbf{n}} \times \hat{\mathbf{k}}_o\right|} = \hat{\varphi}'_o \tag{B.8a}$$

The local vertical polarization is taken perpendicular to both $\hat{\mathbf{k}}_o$ and $\hat{\mathbf{h}}_o$; thus,

$$\hat{\mathbf{v}}_o = \hat{\mathbf{h}}_o \times \hat{\mathbf{k}}_o = \hat{\boldsymbol{\theta}}'_o \tag{B.8b}$$

Similarly, the local horizontal polarization vector, $\hat{\mathbf{h}}_i$, for incident wave is taken parallel to the local ground $(x'y')$ plane. More specifically,

$$\hat{\mathbf{h}}_i = \frac{\hat{\mathbf{k}}_i \times \hat{\mathbf{n}}}{\left|\hat{\mathbf{k}}_i \times \hat{\mathbf{n}}\right|} = \hat{\varphi}'_i \tag{B.9a}$$

The local vertical polarization is taken perpendicular to both $\hat{\mathbf{k}}_i$ and $\hat{\mathbf{h}}_i$; thus,

$$\hat{\mathbf{v}}_i = \hat{\mathbf{k}}_i \times \hat{\mathbf{h}}_i = \hat{\boldsymbol{\theta}}'_i \tag{B.9b}$$

APPENDIX C

ANTENNA ROTATION MATRICES

The received field expressions derived in the main text assume knowledge of the relationship between various coordinate systems. In this appendix, we define coordinate transformation between the antenna coordinates and the reference coordinate system. Let us consider the receive antenna here but the same is also applicable to the transmit antenna. The antenna coordinate system is rotated about the z-axis by φ_{0r} in azimuth. Then, the antenna is rotated about the y-axis by $(\pi - \theta_{0r})$ in elevation. The angle φ_{0r} is defined counterclockwise from the x-axis while the angle θ_{0r} is defined counterclockwise from the z-axis. When $\varphi_{0r} = 0$, the antenna is facing East and when $\varphi_{0r} = \pi/2$, the antenna is facing North. The azimuth rotation is performed for aligning the direction of the incident field from a satellite. When $\theta_{0r} = 0$, the antenna is facing nadir and when $\theta_{0r} = \pi$, the antenna is facing zenith. The angle θ_{0r} represents the observation angle. With these rotations, the antenna y_r-axis will always be parallel to the ground (xy plane), so that it can represent the horizontal-polarized port when the antenna is linearly polarized. The rotation matrices from the antenna to the reference system in elevation and azimuth planes are, respectively, given by

$$[A]_{\theta_{0r}} = \begin{bmatrix} -\cos\theta_{0r} & 0 & \sin\theta_{0r} \\ 0 & 1 & 0 \\ -\sin\theta_{0r} & 0 & -\cos\theta_{0r} \end{bmatrix} \tag{C.1a}$$

$$[A]_{\varphi_{0r}} = \begin{bmatrix} \cos\varphi_{0r} & -\sin\varphi_{0r} & 0 \\ \sin\varphi_{0r} & \cos\varphi_{0r} & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{C.1b}$$

Because the rotation matrix is orthogonal, transformation in the reverse direction is simply the transpose of the matrix.

APPENDIX D

OUTER PRODUCT MATRICES AND VECTORS

The outer product is defined for two-element vectors by [68]

$$\underline{P} = \underline{b} \otimes \underline{b}^* = \begin{bmatrix} b_1 \underline{b}^* \\ b_2 \underline{b}^* \end{bmatrix} = \begin{bmatrix} b_1 b_1^* \\ b_1 b_2^* \\ b_2 b_1^* \\ b_2 b_2^* \end{bmatrix} \tag{D.1}$$

and for the two-by-two matrix of a circularly polarized antenna by

$$\underline{\underline{G}} = \underline{\underline{g}} \otimes \underline{\underline{g}}^* = \begin{bmatrix} g_{RR} \underline{\underline{g}}^* & g_{RL} \underline{\underline{g}}^* \\ g_{LR} \underline{\underline{g}}^* & g_{LL} \underline{\underline{g}}^* \end{bmatrix} = \begin{bmatrix} g_{RR} g_{RR}^* & g_{RR} g_{RL}^* & g_{RL} g_{RR}^* & g_{RL} g_{RL}^* \\ g_{RR} g_{LR}^* & g_{RR} g_{LL}^* & g_{RL} g_{LR}^* & g_{RL} g_{LL}^* \\ g_{LR} g_{RR}^* & g_{LR} g_{RL}^* & g_{LL} g_{RR}^* & g_{LL} g_{RL}^* \\ g_{LR} g_{LR}^* & g_{LR} g_{LL}^* & g_{LL} g_{LR}^* & g_{LL} g_{LL}^* \end{bmatrix} \tag{D.2}$$

The elements of the coherency vector and matrix are complex, and it is sometimes desirable to describe the wave by real quantities. The modified Stokes vector is commonly used in microwave remote sensing to describe both amplitude and polarization of the wave [156]. It is a transform of the coherency vector. The modified Stokes vector of the total power is given by

$$\underline{P}_M = \underline{\underline{Q}} \cdot \underline{P} = \begin{bmatrix} |b_1|^2 \\ |b_2|^2 \\ 2 \operatorname{Re} \{b_1 b_2^*\} \\ 2 \operatorname{Im} \{b_1 b_2^*\} \end{bmatrix} \tag{D.3}$$

where

$$\underline{Q} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & -i & -i & 0 \end{bmatrix} \tag{D.4}$$

Similarly, the four-by-four Mueller matrix of a circularly polarized antenna can be defined for the modified Stokes vector (D.5), as shown at the top of this page.

$$\underline{\underline{G}} = \underline{\underline{g}} \otimes \underline{\underline{g}}^* = \begin{bmatrix} g_{RR}\underline{\underline{g}}^* & g_{RL}\underline{\underline{g}}^* \\ g_{LR}\underline{\underline{g}}^* & g_{LL}\underline{\underline{g}}^* \end{bmatrix}$$

$$= \begin{bmatrix} |g_{RR}|^2 & |g_{RL}|^2 & \mathrm{Re}\{g_{RR}g_{RL}^*\} & -\mathrm{Im}\{g_{RR}g_{RL}^*\} \\ |g_{LR}|^2 & |g_{LL}|^2 & \mathrm{Re}\{g_{LL}g_{LR}^*\} & \mathrm{Im}\{g_{LL}g_{LR}^*\} \\ 2\,\mathrm{Re}\{g_{RR}g_{LR}^*\} & 2\,\mathrm{Re}\{g_{LL}g_{RL}^*\} & \mathrm{Re}\{g_{RR}g_{LL}^* - g_{RL}g_{LR}^*\} & -\mathrm{Im}\{g_{RR}g_{LL}^* - g_{RL}g_{LR}^*\} \\ 2\,\mathrm{Im}\{g_{RR}g_{LR}^*\} & -2\,\mathrm{Im}\{g_{LL}g_{RL}^*\} & \mathrm{Im}\{g_{RR}g_{LL}^* + g_{RL}g_{LR}^*\} & \mathrm{Re}\{g_{RR}g_{LL}^* - g_{RL}g_{LR}^*\} \end{bmatrix} \tag{D.5}$$

APPENDIX E

CYGNSS LEVEL 1 DATA

Table E.1: CYGNSS Level 1 data that are used in this study.

brcs	17 $\times$ 11 array of DDM bin bistatic radar cross section, m^2. The specular point is located in DDM bin round(brcs_ddm_sp_bin_delay_row), round(brcs_ddm_sp_bin_dopp_col).
power_analog	17 $\times$ 11 array of DDM bin analog power, Watts. analog_power is the true power that would have been measured by an ideal (analog) power sensor. power_digital is the power measured by the actual 2-bit sensor, which includes quantization effects. power_analog has been corrected for quantization effects.
ddm_snr	DDM signal to noise ratio. $10log(Smax/Navg)$, where $Smax$ is the maximum value (in raw counts) in a single DDM bin and $Navg$ is the the average per-bin raw noise counts. ddm_snr is in dB.
direct_signal_snr	10log(zenith signal power/zenith signal noise) at ddm_timestamp_utc.
sp_lat	SP latitude, in degrees North.
sp_lon	SP longitude, in degrees East.
rx_to_sp_range	The distance between the CYGNSS spacecraft and the specular point, in meters, at ddm_timestamp_utc.
tx_to_sp_range	The distance between the GNSS spacecraft and the specular point, in meters, at ddm_timestamp_utc.
sp_inc_angle	SP incidence angle, in degrees. This is the angle between the line normal to the Earth's surface at the specular point and the line extending from the specular point to the spacecraft.
gps_tx_power_db_w	GPS SV transmit power. Power input to SV Tx antenna, in dBw.
gps_ant_gain_db_i	GPS SV transmit antenna gain. SV antenna gain in the direction of the specular point, in dBi.
sp_rx_gain	Specular point Rx antenna gain. The receive antenna gain in the direction of the specular point, in dBi.
quality_flags	Per-DDM quality flags. 1 indicates presence of condition.

APPENDIX F

RELEVANT PUBLICATIONS

Full list of publications and presentations that are related to this dissertation is as follows:

Refereed Journal Papers:

- **O. Eroglu**, M. Kurum, D. Boyd, and Ali C. Gurbuz, "High spatio-temporal resolution CYGNSS soil moisture estimates using artificial neural networks," *Remote Sensing*, vol. 11, no. 19, 2019, pp. 1-32.
- **O. Eroglu**, M. Kurum, and J. Ball, "Response of GNSS-R on dynamic vegetated terrain conditions," *IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing*, vol. 12, no. 5, 2019, pp. 1599-1611.
- **O. Eroglu**, D. Boyd, and M. Kurum, "SCoBi: A free, open-source, SoOp Coherent Bistatic Scattering Simulator framework," *IEEE Geoscience and Remote Sensing Magazine*, in press.
- M. Kurum, M. Deshpande, A. Joseph, P. E. O'Neill, R. H. Lang, and **O. Eroglu**, "SCoBi-Veg: A generalized bistatic scattering model of reflectometry from vegetation for Signals of Opportunity applications," *IEEE Transaction on Geosciences and Remote Sensing*, vol. 57, no. 2, 2018, pp. 1049-1068.

Conference Papers and Presentations:

- **O. Eroglu**, M. Kurum, D. R. Boyd, A. C. Gurbuz, "Investigations into CYGNSS-based soil moisture retrieval algorithms," The IEEE International Geosciences and Remote Sensing Symposium (IGARSS'19), Valencia, Spain, Jul. 28Aug. 2, 2019.
- D. R. Boyd, M. Kurum, J. L. Garrison, B. Nold, R. Mdrafi, **O. Eroglu**, A. C. Gurbuz, J. Piepmeier, M. S. Vega, and R. Bindlish, "Inversion of simulated and physical soil moisture profiles using multifrequency SoOP-sources," The IEEE International Geosciences and Remote Sensing Symposium (IGARSS'19), Valencia, Spain, Jul. 28Aug. 2, 2019. (Invited)
- D. R. Boyd, **O. Eroglu**, M.Kurum, "Demo of open-sourced SCoBi modeling simulation tool," Specialist Meeting on Reflectometry using GNSS and other Signals of Opportunity (GNSS+R'19), Benevento, Italy, May. 2022, 2019.
- **O. Eroglu**, D. R. Boyd, A. C. Gurbuz, and M. Kurum, "Relating CYGNSS observations to soil moisture variations during the 2018 hurricane season," Presented at: The USNC/URSI Radio Science Meeting, Boulder, CO (USA), Jan. 912, 2019.
- D. R. Boyd, M. Kurum, **O. Eroglu**, A. C. Gurbuz, J. L. Garrison, B. Nold, J. Piepmeier, M. S. Vega, and R. Bindlish, "Investigation of root-zone soil moisture profile sensitivity to multiple Signal of Opportunity sources," Presented at: The USNC/URSI Radio Science Meeting, Boulder, CO (USA), Jan. 912, 2019.

- M. Kurum and **O. Eroglu**, "On the feasibility of environmental monitoring via a smartphone GNSS receiver," Presented at: The Institute of Navigation (ION) Global Navigation Satellite System (GNSS) Meeting (ION GNSS+'18), Miami, FL, Sep. 2428, 2018.

- **O. Eroglu**, M. Kurum, and J. Ball, "Preliminary results of a GNSS-R simulation to sense canopy parameters," *In: Proceedings of International Conference on Electromagnetics in Advanced Applications (ICEAA'18)*, Cartagena de Indias, Colombia, Sep. 1014, 2018.

- **O. Eroglu**, D. Boyd, M. Kurum, "Open-sourcing of a SoOp simulator with bistatic vegetation scattering model," *In: Proceedings of the IEEE International Geosciences and Remote Sensing Symposium (IGARSS'18)*, Valencia, Spain, Jul. 2327, 2018.

- **O. Eroglu** and M. Kurum, "Could GNSS-reflectometry sense corn growth stages?," Presented at: The USNC/URSI Radio Science Meeting, Boulder, CO (USA), Jan. 47, 2018.

- M. Kurum, M. Deshpande, A. Joseph, P. E. O'Neill, R. H. Lang, and **O. Eroglu**, "Development of a coherent bistatic vegetation model for Signal of Opportunity applications at VHF/UHF bands," *In: Proceedings of the IEEE International Geosciences and Remote Sensing Symposium (IGARSS'17)*, 2017, pp. 4894-4896.

- M. Kurum, M. Deshpande, A. Joseph, P. E. O'Neill, R. H. Lang, and **O. Eroglu**, "A coherent bistatic vegetation model for Signal of Opportunity land applications: Preliminary simulation results at VHF-UHF Bands," Presented at: Specialist Meeting on Reflectometry using GNSS and other Signals of Opportunity (GNSS+R'17), Ann Arbor, MI (US), May. 2325, 2017.